ON ALGEBRAIC ESTIMATION AND SYSTEMS WITH GRADED POLYNOMIAL STRUCTURE

To Liz

ON ALGEBRAIC ESTIMATION AND SYSTEMS WITH GRADED POLYNOMIAL STRUCTURE

by

PETER CHARLES COLLINGWOOD, B.Sc. (WARWICK, 1978)

SUBMITTED IN SATISFACTION OF THE REQUIREMENTS
FOR THE DEGREE OF Ph.D. TO THE

CONTROL THEORY CENTRE
UNIVERSITY OF WARWICK

FEBRUARY 1985

SUMMARY

In the first half of this thesis the algebraic properties of a class of minimal, polynomial systems on $\mathbb{R}^n$ are considered. Of particular interest in the sequel are the results that

(i) a tensor algebra generated by the observation space and strong accessibility algebra is equal to the Lie algebra of polynomial vector fields on $\mathbb{R}^n$

and (ii) the observation algebra of such a system is equal to the ring of polynomial functions on $\mathbb{R}^n$.

The former result is proved directly, but to establish the second we construct a canonical form for which the claim is trivial, the general case then following from the properties of the diffeomorphism relating the two realisations. It is also shown that, as a consequence of the structure of the observation space, any system in the class considered has a finite Volterra series solution, thereby showing that the canonical form developed is dual to that of Crouch.

The second part of the work is devoted to the algebraic aspects of nonlinear filtering. The fundamental question that this 'algebraic estimation theory' seeks to answer is the existence of a homomorphism between a Lie algebra Λ of differential operators and a Lie algebra of vector fields. By restricting Λ to be finite dimensional we obtain a restrictive condition on the system generating Λ. Results of Ocone and Hijab are extended and connections with the work of Omori and de la Harpe established thus showing Λ seldom has a Banach structure. Finally, using an observability condition, we develop a further canonical form and thus define a class of systems for which Λ is isomorphic to the Weyl algebra on n-generators and hence cannot satisfy the above homomorphism principle.

CONTENTS

SUMMARY: (ii)
ACKNOWLEDGEMENTS: (iv)
INTRODUCTION: (v)

I: NONLINEAR SYSTEMS AND GRADED POLYNOMIAL STRUCTURES 1

1.1. Polynomials, vector fields and one forms 2

1.2. Nonlinear systems: theory and local structure 11

1.3. Graded polynomial systems 19

II: CANONICAL REPRESENTATIONS OF GRADED POLYNOMIAL FORMS 32

2.1. The graded observable polynomial form 33

2.2. Finite Volterra Series and g.p.f's 42

III: ALGEBRAIC ESTIMATION 52

3.1. Finite dimensional estimation algebras 54

3.2. Examples 75

IV: AN EXAMPLE OF HAZEWINKEL AND MARCUS 88

4.1. Graded polynomial forms, algebraic estimation and $W_{\bar{n}}$. 90

4.2. Drift independent observability 99

4.3. The estimation algebra for a class of d.i.o. systems 110

BIBLIOGRAPHY 119

Acknowledgements

The list of people to whom I owe thanks for assistance and inspiration in the completion of this thesis, whilst not endless, is long. I would like to express particular gratitude to my supervisor Dr. P.E. Crouch, for showing remarkable fortitude and stamina, to Dr. S.A. Billings for allowing me the time to work on the manuscript, and to Dr. S.P. Banks for bearing up under my persistent and often trivial questioning. I derived great moral support from several other sources including the remaining members of both the Control Theory Centre and the Department of Control Engineering at Sheffield University, the (~~English~~) British Nonlinear Systems Group (especially its 'founding father' and driving force, Dr. 'Hercules' Irving) and last, but not least, the Dark Peak Fell Runners.

Finally, my unending gratitude and admiration goes out to Madeleine Floy for typing the thesis with such skill and patience and also for only swearing when I was out of earshot.

INTRODUCTION

The breadth and wealth of mathematics used in the attempt to analyse (and derive accurate representations of) nonlinear phenomena makes working in the field a veritable fiddlers paradise. Within the confines of Systems Theory this observation is particularly true for the specific problem of constructing recursive estimators of a stochastic diffusion process. Following Kalman and Bucy's pioneering efforts in the case that the state is generated as the solution of a linear system and the recognition of the fundamental rôle played by the innovations process, the rigours of Martingale Theory have been successfully applied; the major achievement of this approach being, undoubtedly, the stochastic differential equation (s.d.e.) for the conditional statistic as derived in Fujisaki, Kallianpur and Kunita [1]. Whilst, in a sense, giving a complete solution to the question of the existence of statistics of the state process, from a practical point of view several obstacles remain, not least the complexity of the systems derived and their non-recursive nature.

In the attempt to overcome these difficulties a comparatively new approach to filtering drawing on the ideas of the differential geometric theory of nonlinear control systems has been developed in which the probabilistic features of the problem are played down. Instead, by using elements from Differential Geometry, Functional Analysis and Lie Algebras a theory has been constructed, giving an algebraic necessary criterion for the existence of 'readily computable' statistics, in which a homomorphism between a Lie algebra, Λ, of differential operators on $\mathbb{R}^n$ and a Lie algebra of vector fields is sought. It is this 'fundamental question of algebraic estimation' which forms the central theme of this thesis and, in particular, that of Chapters III and IV.

There are two immediately obvious ways to construct a general theory on the basis of a necessary condition namely by classifying those objects which either do or do not satisfy the criterion. In the present context it is classical that if Λ is finite dimensional then it is isomorphic to a Lie algebra of matrices and hence can be identified as a Lie algebra of linear vector fields. Thus it is first natural to ask if there are any classes of systems (other than linear) for which Λ, the so-called Estimation Algebra, is finite dimensional and, following a deeper exposition of the ideas behind algebraic estimation, it is this aspect of the problem to which the rest of Chapter III is devoted. As we shall see, it is possible to derive a fairly restrictive condition on the types of system exhibiting this behaviour - essentially the output must be 'quadratic' along trajectories of the input vector field. Having established that finite dimensionality is rare we extend similar results of Ocone and Hijab. In particular, we offer two generalisations of the relationship between the output and the input vector field to the case that the noise entering the system is m-dimensional. Also considered is the interesting situation that noisy observations are made of a deterministic control system with random initial condition, showing that Λ is finite dimensional iff the system has a bilinear realisation. We finish the chapter by discussing some results of Omori and de la Harpe which suggest that not only does the estimation algebra seldom have finite dimension but that it is also unlikely to have a Banach structure, once again highlighting the complexity of the nonlinear filtering problem.

In contrast to these arguments, Chapter IV is devoted to describing a class of systems for which the estimation algebra is isomorphic to the Weyl algebra W_n of all differential operators on $\mathbb{R}^n$ with polynomial coefficients. As Marcus and Hazewinkel have pointed out this suggests that such a system cannot have any finite dimensionally computable (f.d.c.)

statistics since there can be no non-trivial homomorphisms between W_n and a Lie algebra of vector fields. To achieve our construction we first introduce the concept of drift independent observability, a dual notion to the input independent observability discussed by Gauthier, Bornard and Nijmeier, which allows us to obtain a canonical form for this class of systems. By appealing to the results of Chapters I and II we can then reach our desired conclusion by assuming that the system in question also has a particular polynomial structure (an obvious necessary condition) and that certain generators have a-priori been established as elements of Λ.

From this brief description, it is clear that the early part of the thesis was inspired to a large extent by the calculations of the final chapter. However, it is of strong independent interest since it provides an algebraic analysis, revealing a rich structure, of a generic class of non-trivial systems. We begin Chapter I with a brief survey of the theory of graded vector spaces and introduce some of the basic terminology used throughout the thesis. Our investigations start then in §1.2 with a discussion of the local structure of minimal linear analytic systems. It is well-known that controllability and observability of such systems are determined by the "transitivity" properties of certain associated Lie algebras ($\mathscr{L}$ or $\mathscr{S}$) and the observation space $\mathscr{H}$; in particular we require that the (co) distributions on the state space determined by $\mathscr{L}$ or $\mathscr{S}$ and $\mathscr{H}$ should contain a basis for each fibre of the relevant bundle. Thus, it is natural to expect that locally we can find a description of the system for which $\mathscr{L}$, $\mathscr{S}$ or $\mathscr{H}$ contain the corresponding coordinates. This indeed turns out to be true for $\mathscr{H}$, but we also show that the dual result for the vector fields is not. However, by extending the base ring of $\mathscr{S}$ from $\mathbb{R}$ to $\mathbb{R}[x_1,\ldots,x_n]$ we find that any minimal system in graded polynomial form (g.p.f.) possesses this coordinate canonicality property globally. (In fact, we show that the module thus generated is identically the space of polynomial vector fields on $\mathbb{R}^n$).

The primary objective of Chapter II is to obtain a dual to this result, namely that for minimal systems in g.p.f. the observation algebra is the ring of polynomial functions on $\mathbb{R}^n$. We achieve this aim by constructing a global canonical form for which $\mathscr{H}$ always contains the coordinate functions and therefore trivially satisfies $\mathscr{H}_A \equiv \mathbb{R}[x_1,\dots,x_n]$. The general case then follows immediately from a further result of the previous chapter showing that the system diffeomorphism between two minimal g.p.f's is polynomial with polynomial inverse. In the final section of this chapter we discuss an algebraic characterisation of systems with finite Volterra series showing that this class coincides with the g.p. forms and moreover, that the algorithm presented by Crouch for the minimal realisation of such f.v.s. is dual to the construction given here.

For the most part it is hoped that this thesis is self-contained. However, at least a nodding aquaintance with the basic elements and notations from differential geometry, functional analysis, Lie algebras and nonlinear systems theory would prove useful.

CHAPTER I: NONLINEAR SYSTEMS AND GRADED POLYNOMIAL STRUCTURES

The systematic study of nonlinear systems in their most general form, assuming only sufficient regularity and structure to ensure the equations are well-defined, can at best produce only limited results. Whilst of obvious fundamental importance and interest, these theorems tend to be of a local nature and it is only rarely that global implications can be made, usually at the expense of further constraints. Since the primary concern of Systems Theory is the prediction of global behaviour, this is a very serious restriction and for this reason, we are led to question the existence of a class of systems having enough structure to allow strong analysis but which are not on the other hand, too pathological or trivial.

In this chapter we present a step in this direction by considering the properties of a class of systems which, although they have an intuitively natural form, have not formed the basis for any previous consistent analysis. Moreover, it is shown that there is associated with each such system a very rich algebraic structure, some of whose implications are exploited in later chapters but which may also prove to have important consequences in control design and other, more practical, aspects of systems theory. Further properties, and indeed their relationship to the general scheme of nonlinear systems, are established in the next chapter, but here we concentrate on those aspects dealing with controllability and diffeomorphisms between minimal representations. We begin by surveying and establishing most of the notation and concepts, used throughout this thesis in §1.1. In particular, a generalised form of the notion of a homogeneous polynomial is presented and the induced structure on the space of polynomial vector fields is studied. In the second section, the local structure of nonlinear systems is examined, particularly with reference to coordinate canonicality. It will be seen that 'controllability' and 'observability' of a nonlinear system can be determined by the calculation of a (Lie) algebra of vector fields and a vector space of functions: it is natural to ask if the resulting system algebra or observation space contain the relevant coordinates used in these computations. If either of these circumstances apply the system is said to be controllably (resp. observably) coordinate canonical. It is shown that any minimal system will be o.c.c. but may not be controllably so. Finally, in §1.3, the class of systems to be studied is introduced, namely, the graded polynomial forms.

§1.1. Polynomials, Vector Fields and One Forms

This section is primarily concerned with notation and the consequences of a generalised definition of homogeneity of polynomials on the subsequent induced structure of the spaces of polynomial vector fields and one forms. For further details of the material presented here,we refer to Goodman [1]

We begin by recalling that a polynomial $\phi : \mathbb{R}^n \to \mathbb{R}$ is said to be homogeneous of degree k if, for any $s \in \mathbb{R}^+$, $\phi(sx) = s^k \phi(x)$. Standard examples of such functions are constructed by considering a set of coordinates $\{x_1,\ldots,x_n\}$ for $\mathbb{R}^n$ and then letting ϕ be a finite linear combination of elements of the form $x_1^{\alpha_1} \ldots . x_n^{\alpha_n}$ with $|\alpha| \triangleq \alpha_1 + .. + \alpha_n = k$. This concept can be generalised in the following manner (although the details given here are for $\mathbb{R}^n$, we remark that the analysis is equally valid for any finite dimensional vector space). For a given set of integers $n_1,\ldots,n_p$ such that $\sum_{i=1}^{p} n_i = n$ and $n_i \geq 0$ we can decompose $\mathbb{R}^n$ into a direct sum of p subspaces, $\bigoplus_{i=1}^{p} \mathbb{R}^{n_i}$. Any element $\underline{x} \in \mathbb{R}^n$ can then be written, equivalently, as either $\underline{x} = \underline{x}_1 \oplus \ldots . \oplus \underline{x}_p$ or $\underline{x} = (\underline{x}_1,\ldots,\underline{x}_p)$ with each component $\underline{x}_i \in \mathbb{R}^{n_i}$, $1 \leq i \leq p$, in turn having components $(x_i^1,\ldots,x_i^{n_i})$. Next, let $\{\delta_t \, ; \, t > 0\}$ be the group of dilations of $\mathbb{R}^n$ satisfying

$$\delta_t(x) = \bigoplus_{i=1}^{p} t^i \underline{x}_i = (t\underline{x}_1, \ldots, t^p \underline{x}_p)$$

$$\delta_t \delta_s = \delta_{ts}$$

(so that each δ_t is a diffeomorphism with $\delta_t^{-1} = \delta_{1/t}$). The pair $(\mathbb{R}^n = \oplus \mathbb{R}^{n_i}, \delta_t)$ is said to be a graded vector space of degree p. We can now define a sequence of subspaces of $\mathbb{R}[x_1, \ldots, x_n]$, the algebra of real valued polynomial functions on $\mathbb{R}^n$, by setting

$$H^k = \{\phi \; ; \; \phi \circ \delta_t = t^k \phi\} \qquad k \geq 0$$

Clearly, if the gradation of $\mathbb{R}^n$ is of degree 1, then the spaces H^k will coincide with the standard homogeneous polynomials described above. For this reason, H^k is defined to be the space of homogeneous polynomials of weight k.

It is also straightforward to construct examples of such polynomials in the general case by considering an ordered basis $\{x_1^1, \ldots, x_1^{n_1}, x_2^1, \ldots, x_2^{n_2}, \ldots\ldots x_p^{n_p}\}$ for $\mathbb{R}^n$. Then any finite combination of elements of the form

$$x^{\alpha} = (x_1^1)^{\alpha_{11}} \ldots (x_p^{n_p})^{\alpha_{pn_p}} = \underline{x}_1^{\alpha_1} \ldots \underline{x}_p^{\alpha_p}, \; \alpha_i = (\alpha_{i1}, \ldots, \alpha_{in_i})$$

such that $w(\alpha) \overset{\Delta}{=} \sum_{k=1}^{p} k|\alpha_k| = m$, is an element of H^m. In more concrete terms, suppose that $\mathbb{R}^3$ is decomposed as $\mathbb{R} \oplus \mathbb{R}^2$ with $\underline{x}_1 = x_1$ and $\underline{x}_2 = \binom{x_2}{x_3}$. Then it is readily seen that

$$H^0 = \mathbb{R}, \quad H^1 = Sp\{x_1\}, \quad H^2 = Sp\{x_1^2, x_2, x_3\}$$

$$H^3 = Sp\{x_1^3, x_1x_2, x_1x_3\} \qquad H^4 = Sp\{x_1^4, x_1^2x_2, x_1^2x_3, x_2x_3, x_2^2, x_3^2\}$$

etc.

where $Sp\{\cdot\}$ denotes the linear span over $\mathbb{R}$ of the elements enclosed in $\{\cdot\}$. Clearly, different sets of polynomials will be obtained for different decompositions of $\mathbb{R}^3$. We also remark that the definition allows for some of the subspaces to be trivial. For instance, $\mathbb{R}^3$ can be written as

$\mathbb{R} \oplus \mathbb{R}^0 \oplus \ldots . \oplus \mathbb{R}^0 \oplus \mathbb{R}^2 = \bigoplus_{i=1}^{p+1} \mathbb{R}^{n_i}$ with $n_1 = 1, n_{p+1} = 2$ and $n_i = 0$ for $2 \leq i \leq p$. In this case, we have, for example $H^0 = \mathbb{R}$ & $H^k = Sp\{x_1^k\}$ for $2 \leq k \leq p$.

Many of the standard results on the algebraic structure of the polynomial algebra can be reinterpreted in the light of the above definitions. Foremost amongst these, for our purposes, is the construction of a filtration $\{Q^m : m \geq 0\}$ of $\mathbb{R}[x_1,\ldots,x_n]$ obtained by setting

$$Q^m = \bigoplus_{k=0}^{m} H^k$$

and satisfying

(i) $\mathbb{R} = Q^0 \subset Q^1 \subset \ldots \subset Q^m \subset \ldots$

(ii) $\bigcup_{m \geq 0} Q^m = \mathbb{R}[x_1,\ldots,x_n]$

(iii) $Q^m \otimes Q^n \subset Q^{m+n}$

Again, in analogy with the standard definitions, Q^m is defined to be the space of polynomials of weight $\leq m$.

Also of importance in our analysis will be the graded form of the Taylors series. However to introduce this, we also need the concept of a dilation homogeneous norm on $\mathbb{R}^n$.

DEFINITION 1.1.1(Goodman [1])

A dilation homogeneous norm on $\mathbb{R}^n$ is a continuous function $x \to ||x||_\delta$ taking values in $\mathbb{R}^+$ and satisfying

(i) $||x||_\delta = 0 \iff x = 0$

(ii) $||\delta_t x||_\delta = t||x||_\delta \qquad t > 0.$ □

Examples of such functions are given by the following generalisations of the usual p-norms

$$||x||_{p,\delta} = \left(\sum_{i=1}^{r} \sum_{j=1}^{n_i} |x_i^j|^{p/i} \right)^{1/p} \qquad \text{for a gradation of degree } r$$

$$||x||_{\infty,\delta} = \max_{i,j} |x_i^j|^{1/i} \qquad 1 \leq i \leq r,\ 1 \leq j \leq n_i$$

where $|.|$ is the modulus on $\mathbb{R}$. It can be shown that all homogeneous norms on $\mathbb{R}^n$ are equivalent. Moreover, if $p \varepsilon H^k$ and $x \varepsilon \mathbb{R}^n \backslash \{0\}$, with $||x||_\delta^{-1} \triangleq t_o$, then

$$|p(\delta_{t_o} x)| = |t_o^k p(x)| = t_o^k\ |p(x)|$$

and, since $||\delta_{t_o}(x)||_\delta = t_o||x||_\delta = 1$, it follows that

$$|p(x)| \leq \max_{||u||_\delta = 1} (|p(u)|)\ ||x||_\delta^k .$$

Conversely, if $p \varepsilon Q^m$ and satisfies $|p(x)| \leq M||x||_\delta^k$, then p must be an element of H^k. For, we can write $p = \sum_{\ell=0}^{m} p_\ell$ with $p_\ell \varepsilon H^\ell$ so that $p \circ \delta_t = \sum_{\ell=0}^{m} t^\ell p_\ell$. But, by assumption

$$\begin{aligned} |p \circ \delta_t(x)| &\leq M\ ||\delta_t(x)||_\delta^k \\ &\leq M t^k ||x||_\delta^k \end{aligned}$$

so that $p \circ \delta_t = O(t^k)$. By letting $t \to \infty$ we see that $p_\ell = 0$ for $\ell > k$ and, similarly, letting $t \to 0$ we find $p_\ell = 0$ for $\ell < k$.

The Taylors series expansion of a C^∞ function $\phi : \mathbb{R}^N \to \mathbb{R}$ also has a convenient description in terms of these concepts. Clearly, about $x = 0$ say, we can write

$$\phi(x) = p_n(x) + r_{n+1}(x)$$

where $n \geq 0$ & $p_n \varepsilon Q^n$ (p_n is the polynomial formed by the elements of weight $\leq n$ in the expansion of ϕ). Then $r_{n+1}(x)$ will be a sum (possibly infinite) of elements of H^m for $m \geq n+1$, so

$$\begin{aligned} r_{n+1}(\delta_t x) &= \sum_{m=n+1}^{\infty} \hat{r}_m(\delta_t x) \\ &= \sum t^m \hat{r}_m(x) \\ &= O(t^{n+1}) \qquad\qquad \forall t. \end{aligned}$$

In particular, we choose $t = ||x||_\delta$ and thus see that $\forall \phi \varepsilon C^\infty(\mathbb{R}^N)$ there is a where $|.|$ is the modulus on $\mathbb{R}$. It can be shown that all homogeneous norms on $\mathbb{R}^n$ are equivalent. Moreover, if $p \in H^k$ and $x \in \mathbb{R}^n \backslash \{0\}$, with $||x||_\delta^{-1} \triangleq t_o$, then

$$|p(\delta_{t_o} x)| = |t_o^k p(x)| = t_o^k \, |p(x)|$$

and, since $||\delta_{t_o}(x)||_\delta = t_o ||x||_\delta = 1$, it follows that

$$|p(x)| \leq \max_{||u||_\delta = 1} (|p(u)|) \, ||x||_\delta^k .$$

Conversely, if $p \in Q^m$ and satisfies $|p(x)| \leq M||x||_\delta^k$, then p must be an element of H^k. For, we can write $p = \sum_{\ell=0}^{m} p_\ell$ with $p_\ell \in H^\ell$ so that $p \circ \delta_t = \sum_{\ell=0}^{m} t^\ell p_\ell$. But, by assumption

$$\begin{aligned} |p \circ \delta_t(x)| &\leq M \, ||\delta_t(x)||_\delta^k \\ &\leq M t^k ||x||_\delta^k \end{aligned}$$

so that $p \circ \delta_t = O(t^k)$. By letting $t \to \infty$ we see that $p_\ell = 0$ for $\ell > k$ and, similarly, letting $t \to 0$ we find $p_\ell = 0$ for $\ell < k$.

The Taylors series expansion of a C^∞ function $\phi : \mathbb{R}^n \to \mathbb{R}$ also has a convenient description in terms of these concepts. Clearly, about $x = 0$ say, we can write

$$\phi(x) = p_n(x) + r_{n+1}(x)$$

where $n \geq 0$ & $p_n \in Q^n$ (p_n is the polynomial formed by the elements of weight $\leq n$ in the expansion of ϕ). Then $r_{n+1}(x)$ will be a sum (possibly infinite) of elements of H^m for $m \geq n+1$, so

$$\begin{aligned} r_{n+1}(\delta_t x) &= \sum_{m=n+1}^{\infty} \hat{r}_m(\delta_t x) \\ &= \sum t^m \hat{r}_m(x) \\ &= O(t^{n+1}) \qquad \forall t. \end{aligned}$$

In particular, we choose $t = ||x||_\delta$ and thus see that $\forall \phi \in C^\infty(\mathbb{R}^n)$ there is a

$p_n \epsilon Q^n$ and $C \epsilon \mathbb{R}^+$ satisfying

$$|p(x) - \phi(x)| \leq C \, ||x||_\delta^{n+1}$$

The final space of functions we need to introduce is $C^\omega(\mathbb{R}^n)$; the analytic functions on $\mathbb{R}^n$. By defining

$$C_m = \{\phi \epsilon C^\omega(\mathbb{R}^n) \ ; \ f(x) = 0(||x||_\delta^m) \text{ near } 0\}$$

we obtain a filtration on $C^\omega(\mathbb{R}^n)$ with

$$C^\omega(\mathbb{R}^n) = C_o \supset C_1 \supset \ldots.$$

$$C_m \otimes C_n \subseteq C_{m+n}$$

Now, if $\phi \epsilon C_m$, by the graded Taylors series we have that, for some $r \epsilon C_{m+1}$

$$\phi = p + r$$

However, $C_{m+1} \subset C_m$, so $\phi - r \epsilon C_m$. This implies that $p \epsilon C_m$ and hence, by the previous analysis, $p \epsilon H^m$. In other words, we obtain the decompositions

$$C_m = H^m \oplus C_{m+1}$$

and

$$C_o = Q^m \oplus C_{m+1} \qquad \forall \ m \geq 0.$$

The above structure we have introduced on several function spaces induces similar structure on the spaces of vector fields and one forms on $\mathbb{R}^n$. Before describing these filtrations, we present some nomenclature from differential geometry which both simplifies some of the expressions to be derived and which will also be used throughout this thesis.

So, let M^n be a smooth n-dimensional manifold (for convenience, 'smooth' is usually taken to be C^∞). Then TM and T^*M will denote the tangent and cotangent bundles of M respectively whilst $\Gamma^r(TM)$ resp. $\Gamma^r(T^*M)$ are the corresponding spaces of C^r vector fields and one forms on M for $r=0,1,\ldots.,\infty, \omega$. Typically, given a coordinate chart (U,x) about $x_o \epsilon M$, $X \epsilon \Gamma^r(TM)$ and $\omega \epsilon \Gamma^r(T^*M)$ will be written

$$X(x) = \sum_{i=1}^{n} X_i(x) \frac{\partial}{\partial x_i} \quad \text{with } X_i \varepsilon C^r(U) \text{ and } X(x) \varepsilon T_x M$$

$$\omega(x) = \sum_{i=1}^{n} \omega_i(x) dx^i. \quad \text{with } \omega_i \varepsilon C^r(U) \text{ and } \omega(x) \varepsilon T_x^* M.$$

A given vector field, X, can act in two ways to produce functions namely by the processes of contraction (of one forms) and Lie differentiation (of functions). More specifically, we define operators L_X, which maps

$C^\infty(M) \to C^\infty(M)$, and $i_X : \Gamma^\infty(T^* M) \to C^\infty(M)$ by

$$L_X(\phi)(x) = \frac{d}{dt} (\phi(\gamma_t(x)) \Big|_{t=0} \qquad \text{with } \phi \varepsilon C^\infty(M),\ x \varepsilon M$$

$$i_X(\omega)(x) = \omega_x(X(x)) \qquad \text{with } \omega \varepsilon \Gamma^\infty(T^* M),\ x \varepsilon M.$$

Where $\gamma_t(x)$ is the trajectory of X satisfying $\gamma_o(x) = x$. In a coordinate neighbourhood we then find

$$L_X(\phi)(x) = \sum_{i=1}^{n} X_i(x) \frac{\partial \phi}{\partial x_i} \Big|_x$$

$$i_X(\omega)(x) = \sum_{i=1}^{n} \omega_i(x) X_i(x) = < \omega(x), X(x) >$$

Finally, if Φ: $M \to N$ is a C^∞ diffeomorphism (or, with suitable modifications, a local diffeomorphism), then Φ induces two maps; $\Phi_* : \Gamma^\infty(TM) \to \Gamma^\infty(TN)$ and $\Phi^* : \Gamma^\infty(T^* N) \to \Gamma^\infty(T^* M)$. The formal definitions of these functions are

$$L_{\Phi_* X}(\phi)(n) = L_X(\phi \circ \Phi)(\Phi^{-1}(n)) \qquad \text{for } \phi \varepsilon C^\infty(N), \text{ and } n \varepsilon N$$

$$i_X(\Phi^* \omega)(m) = i_{\Phi_* X}(\omega)(\Phi(m)) \qquad \text{for } \omega \varepsilon \Gamma^\infty(T^* N), \text{and } m \varepsilon M$$

whilst in coordinates we have that, locally,

$$\text{if } (\Phi_* X)(n) = \Sigma Y_i(n) \frac{\partial}{\partial y_i} \text{ then } Y_i(n) = L_X(\Phi_i)(\Phi^{-1}(n))$$

and

$$\text{if } (\Phi^* \omega)(m) = \Sigma \eta_i(m) dx^i \text{ then } \eta_i(m) = \sum_{j=1}^{n} \omega_i(\Phi(m)) \frac{\partial \Phi_i}{\partial x_j} \Big|_m .$$

Now, as before, we suppose that $\mathbb{R}^n$ is graded of degree p. We denote by $D_1(\mathbb{R}^n)$ (resp. $D^1\mathbb{R}^n$)) the vector space of vector fields (resp. one forms) on $\mathbb{R}^n$ which have polynomial coefficients when expressed in terms of the standard coordinates used to define the gradation. Then it is readily seen that both $D_1(\mathbb{R}^n)$ and $D^1(\mathbb{R}^n)$ are filtered vector spaces (with $D_1(\mathbb{R}^n)$ a filtered Lie algebra) with filtrations

$$\{0\}=V_p \subset V_{p-1} \subset .. \subset V_1 \subset V_0 \subset V_{-1} \subset .....$$

$$\{0\} = W_o \subset W_1 \subset .. \subset W_k \subset .....$$

where

$$V_k = \{X \varepsilon D_1(\mathbb{R}^n)\ ;\ L_X(Q^m) \subset Q^{m-k},\ \forall m \geq 0\}$$

$$W_k = \{\omega \varepsilon D^1(\mathbb{R}^n)\ ;\ i_X(\omega) \varepsilon Q^{k-m},\ \forall X \varepsilon V_m\}$$

These subspaces can be easily characterised in terms of the degree of their coefficients as follows.

<u>THEOREM 1.1.2</u>

a) With the above sequence of subspaces, $D_1(\mathbb{R}^n)$ is a filtered Lie algebra

b) $\forall\ k \leq p,\ \ell \geq 0$

(i) $V_k = \bigoplus_{j=1}^{p} Q^{j-k} \otimes \Delta_j$

(ii) $W_\ell = \bigoplus_{j=1}^{p} Q^{\ell-j} \otimes \Delta^j$

where $\Delta_j = \mathrm{Sp}\{ \frac{\partial}{\partial x_j^1}, \ldots, \frac{\partial}{\partial x_j^{n_j}} \}$ and $\Delta^j = \mathrm{Sp}\{dx_j^1, \ldots, dx_j^{n_j}\}$.

<u>Proof</u>

a) By definition, we need only show that $[V_j, V_k] \subset V_{j+k},\ \forall\ j,k \leq p$.

But, $\forall m \geq 0$, $V_j(Q^m) \subset Q^{m-j}$. So, $\forall m$

$$[V_j,\ V_k](Q^m) \subset V_j\ V_k(Q^m) - V_k\ V_j(Q^m) \subset Q^{m-(k+j)}$$

thus proving the claim

b) (i) Let $X \varepsilon V_k$, with coordinate description $\Sigma X_i(x) \frac{\partial}{\partial x_i}$.

Then, since each coordinate function $x \mapsto x_j$ is an element of Q^{m_j} for some $1 \leq m_j \leq p$, we see that

$$L_{\underline{X}}(x_j) = X_j(x) \in Q^{m_j - k}$$

Thus, $V_k \subset \bigoplus_{j=1}^{p} Q^{j-k} \oplus \Delta_j$.

Conversely, note that the integral curve, $\gamma_{jk}(t)(y)$, of the constant vector field $\frac{\partial}{\partial x_j^k}$ is given by

$$\gamma_{jk}(t)(y) = (\underline{y}_1, \ldots, \underline{y}_{j-1}, y_j^1, \ldots, y_j^k + t, \ldots y_j^{n_j}, \underline{y}_{j+1}, \ldots, \underline{y}_p)^\tau$$

and, hence,

$$\gamma_{jk}(t)(\delta_s y) = \delta_s(\gamma_{jk}(\tfrac{t}{s^j})y).$$

So, $\forall\, \phi \in H^m$, we have

$$\begin{aligned} L_{\frac{\partial}{\partial x_j^k}}(\phi)(\delta_s y) &= \frac{d}{dt}\ \phi(\gamma_{jk}(t)(\delta_s y))\Big|_{t=0} \\ &= s^m \frac{d}{dt}\phi(\gamma_{jk}(\tfrac{t}{s^j})y)\Big|_{t=0} \\ &= s^{m-j}\, L_{\frac{\partial}{\partial x_j^k}}(\phi)(y) \end{aligned}$$

Thus, $L_{\frac{\partial}{\partial x_j^k}}(H^m) \subset H^{m-j}$. Consequently, $\frac{\partial}{\partial x_j^k} \in V_j$ by definition of Q^m, or in other words, $\Delta_j \subset V_j$ The result then follows from the filtration properties of the sequence $\{Q^m ;\ m \geq 0\}$. (ii) is proved similarly. First define

$$\hat{W}_m = \bigoplus_{j=1}^{p} Q^{m-j} \oplus \Delta^j \qquad \subset \Gamma^\omega(T^*\mathbb{R}^n)$$

Then, from part (i), we see that

$$\begin{aligned} i_{V_k}(\hat{W}_m) &\subset \bigoplus_{j=1}^{p} Q^{j-k} \oplus Q^{m-j} \\ &\subset Q^{m-k} \end{aligned}$$

so $\hat{W}_m \subset W_m$. Conversely, we know from the above that $\frac{\partial}{\partial x_j^k} \varepsilon V_j$. Thus, for $\omega \varepsilon W_m$ with $\omega = \sum_{j=1}^{p} \sum_{k=1}^{n_j} \omega_j^k \, dx_j^k$, we see

$$i_{\frac{\partial}{\partial x_j^k}} (\omega) = w_j^k \varepsilon Q^{m-j}$$

Hence $W_m \subset \hat{W}_m$, as required.

$\square$

In a similar fashion, we can also impose a filtration $\{\mathscr{L}_k \; ; \; k \leq p\}$ on $\Gamma^{\omega}(T\mathbb{R}^n)$ by setting

$$\mathscr{L}_k = \{X \varepsilon \Gamma^{\omega}(T\mathbb{R}^n) \; ; \; L_X(C_m) \subset C_{m-k}, \forall m \geq 0\}$$

(by convention we assume $C_m = C_o$ for $m \leq 0$), and, as in Thm. 1.1.2(a), it is easily seen that

$$[\mathscr{L}_k, \mathscr{L}_j] \subset \mathscr{L}_{k+j}$$

$\mathscr{L}_k$ is defined to be the space of (analytic) vector fields of order $\leq k$. From the decomposition $C_m = H^m \oplus C_{m+1}$, we see that $\Delta_j \subset \mathscr{L}_j$. For, by analyticity

$$C_m = \bigoplus_{k=m}^{\infty} H^k$$

and $\Delta_j \subset V_j$. Consequently,

$$\Delta_j(C_m) = \bigoplus_{k=m}^{\infty} \Delta_j(H^k)$$

$$\subset \bigoplus_{k=m}^{\infty} H^{k-j} = C_{m-j}$$

Further, since $\mathbb{R} \subset Q^{\ell}$, $\ell \geq 0$, it follows that $Q^{\ell} \oplus C_m = C_m$. Thus by applying the construction for V_k given in Th$^{\underline{m}}$ 1.1.2 (b) we see that

$$V_k \subset \mathscr{L}_k \qquad \forall \; k \leq p.$$

$\square$

Finally in this section we state two results which are used repeatedly in Chapters 1 and 2.

<u>THEOREM</u> (Palais' Global Inverse Function Theorem (G.I.F.T.) [1]).

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a smooth map. Then f is a smooth diffeomorphism iff

(i) Df_x is invertible $\forall x \in \mathbb{R}^n$

(ii) $||f(x)|| \to \infty$ as $||x|| \to \infty$.

$\square$

THEOREM (see, for instance, Abraham & Marsden [1])

Let $X \in \Gamma^\infty(TM)$ and $\phi \in C^\infty(M)$. Then $L_X(\phi) = 0 \iff \phi$ is constant along trajectories of X

$\square$

§1.2 Nonlinear Systems: Theory and Local Structure

In this section we summarise the main results and constructions of the differential geometric theory of nonlinear systems to be used in the sequel. This has two purposes. Most obviously, it establishes more notation and concepts, however its prime objective is to place in context much of the material presented later. In particular, it motivates the emphasis placed on the algebraic structure theory developed in the next section by highlighting the importance of certain associated algebras.

We shall restrict attention to the class of linear analytic systems, although all of the material presented is equally valid for more general systems, and refer to Sussmann & Jurdjevic [1] and Hermann & Krener [1] for further details. Thus, we shall consider systems described by

$$(1.2.1) \qquad \Sigma \quad \begin{cases} \dot{x} = f(x) + \sum_{i=1}^{m} u_i(t) g_i(x) & u(t) = (u_1(t), \ldots, u_m(t))^T \\ & u(t) \in \Omega \subset \mathbb{R}^m \\ y_j = h_j(x(t)) & 1 \leq j \leq p \end{cases}$$

where $x(t)$ evolves in a finite dimensional, analytic, connected manifold M^n, $f, g_i \in \Gamma^\omega(TM)$ and $h_j \in C^\omega(M)$, $1 \leq i \leq m$, $1 \leq j \leq p$. Moreover, we shall assume that for any input value $u \in \Omega$, the corresponding vector field $f + \Sigma u_i g_i$ is complete ie the corresponding trajectory $t \mapsto x^u(t; t_o, x_o)$, or just $x^u(t)$, satisfying $x^u(t_o) = x_o$, exists $\forall t \in \mathbb{R}$. (We do not define explicitly the class $\mathcal{U}$ of admissible inputs, but remark that, from a technical point of view, it must include the piecewise constant functions and be closed under concatenation). The following definitions are, by now, standard.

DEFINITION 1.2.1. (Sussmann and Jurdjevic [1]

a) The T-reachable set from x_o, $R(x_o,T)$, is defined as

$$R(x_o,T) = \{x\varepsilon M \; ; \; x = x^{u_1}(t_1;t_2,x^{u_2}(t_2;t_3,x^{u_3}(t_3;\ldots\ldots x^{u_n}(t^n;t_o,x_o)\ldots.)$$

$$\forall \; u_i\varepsilon\Omega, t_i\varepsilon\mathbb{R}^+, \text{s.t.} \sum_{i=1}^{n} t_i = T, \; 1 \leqslant i \leqslant n\}$$

b) Σ is accessible, (resp. strongly accessible) if $R(x_o) \triangleq \bigcup_{T>0} R(x_o,T)$ has non-empty interior in M, (resp. $\exists \; T^* > 0$ s.t. $R(x_o,T^*)$ has non-empty interior in M).

□

Accessibility and strong accessibility are natural extensions of the linear concept of controllability (or the requirement that $R(x_o) = M$, $x_o\varepsilon M$). However, whilst it is clear that controllability implies accessibility, further technical hypotheses (for instance on the topological nature of the state space) may be necessary to show that it also implies strong accessibility (Sussmann and Jurdjevic [1], Elliott [1]), and indeed to show that accessibility can be equivalent to strong accessibility. We also remark that if Σ is strongly accessible, then $R(x_o,T)$ has non-empty interior $\forall T > 0$.

Observability for nonlinear systems can be defined in terms of state distinguishability.

DEFINITION 1.2.2. (Hermann and Krener [1])

a) The map $\Sigma_{x_o}(u) \triangleq (h_1(x^u(.;t_o,x_o)),\ldots,h_p(x^u(.;t_o,x_o))^T$ defined on $\mathcal{U} \times M$ is the input-output map of Σ. Two points $x_1, x_2\varepsilon M$ are said to be indistinguishable if $\Sigma_{x_1}(u) = \Sigma_{x_2}(u)$, $\forall u\varepsilon\mathcal{U}$.

b) Σ is observable if no two points of M are indistinguishable. Σ is weakly observable if $\forall x_o\varepsilon M$, $\exists$ a neighbourhood V of x_o in M such that if $x_1\varepsilon V$ and $\Sigma_{x_1}(u) = \Sigma_{x_o}(u)$, $\forall u\varepsilon\mathcal{U}$, then $x_1 = x_o$.

□

As with accessibility, this definition of observability is weaker than that used for linear systems since in this latter case, any input will distinguish between states. However, as in the linear case, if we define a system to be minimal if it is strongly accessible and observable then it is possible to construct minimal realisations of a given (realisable) input-output map. Moreover, any two minimal realisations are related by a unique state space diffeomorphism preserving not only trajectories, but also certain algebraic objects which we now define.

DEFINITION 1.2.3.

a) Let Σ be a linear analytic system as in (1.2.1). The accessibility (resp. strong accessibility) algebra $\mathscr{L}$ (resp $\mathscr{S}$) is defined as the Lie algebra with the following generators

$$\mathscr{L} = \mathscr{L}(\Sigma) \triangleq \{f, g_1, \ldots, g_m\}_{L.A.} \; (\text{resp}\, \mathscr{S} = \mathscr{S}\,(\Sigma) \triangleq \{ad_f^k g_i ; k \geq 0,\ 1 \leq i \leq m\}_{L.A})$$

where $ad_f^0(g_i) = g_i$ and $ad_f^{k+1} g_i = [f, ad_f^k g_i]$. Thus,

$$\mathscr{S} \subset \mathscr{L} \subset \Gamma^\omega(TM)$$

b) With Σ as in part a) the observation space $\mathscr{H} = \mathscr{H}(\Sigma)$ is defined as the smallest subspace of $C^\omega(M)$ and closed under Lie differentiation by the vector fields $f, g_1, \ldots, g_m$ (and, hence, by any element in $\mathscr{L}$). An element $\phi \in \mathscr{H}$ is said to be an observable function.

□

The importance of these spaces is explained by the following theorem, giving algebraic characterisations of accessibility and weak observability.

THEOREM 1.2.4. (Sussmann and Jurdjevic [1], Hermann and Krener [1]

a) A linear analytic system is accessible (resp. strongly acc.) if $\mathscr{L}$ (resp. $\mathscr{S}$) is a transitive Lie algebra (ie spans T_xM at every point $x \in M$). In this case, the group of diffeomorphisms of M generated by the trajectories of the vector fields in $\mathscr{L}$ (or $\mathscr{S}$) acts transitively on M.

b) An accessible linear analytic system is weakly observable if the co-distribution $x \to d\mathscr{H}_x$ has full rank at all points $x \in M$, where $d\mathscr{H} = \{d\phi ; \phi \in \mathscr{H}\} \subset \Gamma^{\omega}(T^*M)$.

□

It turns out that the algebras and observation space of a minimal system are essentially invariants of a given input-output map. Before we prove this, we should first show that, since we are concentrating on the linear-analytic systems this class is closed under the operation of finding a minimal realisation. This is, however, a trivial corollary of the following result.

<u>THEOREM 1.2.5</u>. (Sussman [1])

Let Σ_1 be an analytic, complete system defined on a connected manifold M_1. Then $\exists$ a minimal, analytic complete system Σ_2 evolving on M_2 and an analytic map $\Phi : M_1 \to M_2$ s.t. $\forall x_o \in M_1$ $\Sigma^1_{x_o}(u) = \Sigma^2_{\Phi(x_o)}(u)$ $\forall u \in \mathscr{U}$. Moreover, Φ preserves trajectories. In particular, if Σ_1 is also minimal, then Φ is a diffeomorphism.

□

Now suppose that Σ_1 is linear analytic, with state vector $x_1(t)$ and that Σ_2 is the minimal system realising Σ_1 guaranteed by the above result. Then we have

$$\Phi(x_1^u(t;t_o,x_o)) = x_2^u(t;t_o,\Phi(x_o))$$

where $x_2^u(t)$ is the corresponding trajectory of Σ_2. Differentiating w.r.t.t we find that

$$\begin{aligned} \dot{x}_2 &= F(x_2,u) = \Phi_*(f + \Sigma u_i g_i)(x_2) \\ &= \hat{f}(x_2^u) + \Sigma u^i g_i(x_2^u) \ , \ \hat{f} = \Phi_* f, \ g_i = \Phi_* g_i \ . \end{aligned}$$

Thus, along the orbits of Σ_1 and Σ_2 the dynamics are linear in the input. By analyticity and accessibility, it therefore follows that Σ_2 is also linear analytic.

THEOREM 1.2.6.

Let Σ_1 and Σ_2 be two linear analytic realisations of the same map.

a) If Σ_2 is accessible, then there is a unique linear map $\beta : \mathscr{H}(\Sigma_1) \to \mathscr{H}(\Sigma_2)$ satisfying $\beta(\phi)(x_2^u(t)) = \phi(x_1^u(t))$.

b) If Σ_2 is weakly observable, then there is a unique Lie algebra homomorphism $\gamma : \mathscr{L}(\Sigma_1) \to \mathscr{L}(\Sigma_2)$ satisfying $\gamma(f_1) = f_2$ and $\gamma(g_{1j}) = g_{2j}$.

Proof

a) Define $\hat{\mathscr{H}}(\Sigma_1) = \{\phi_1 \varepsilon \mathscr{H}(\Sigma_1) ; \exists\, \phi_2 \varepsilon \mathscr{H}(\Sigma_2)$ with $\phi_1(x_1^u(t)) = \phi_2(x_2^u(t))\}$

Then $\hat{\mathscr{H}}(\Sigma_1)$ is non-empty since, by definition, it must contain the output functions h_{1j} of Σ_1. Further, if $\phi_1 \varepsilon \hat{\mathscr{H}}(\Sigma_1)$ then

$$\begin{aligned} \frac{d}{dt}\, \phi_1(x_1^u(t) &= L_{f_1}(\phi_1)(x_1^u(t)) + \Sigma u_j L_{g_{1j}}(\phi_1)(x_1^u(t)) \\ &= \frac{d}{dt}\, \phi_2(x_2^u(t)) \qquad \text{for some } \phi_2 \varepsilon \mathscr{H}(\Sigma_2) \\ &= L_{f_2}(\phi_2)(x_2^u(t)) + \Sigma u_j L_{g_{2j}}(\phi_2)(x_2^u(t)). \end{aligned}$$

Since these identities hold $\forall u \varepsilon \mathscr{U}$, it follows that along trajectories

$$L_{f_1}(\phi_1) = L_{f_2}(\phi_2) \qquad L_{g_{1j}}(\phi_1) = L_{g_{2j}}(\phi_2) \qquad 1 \leq j \leq m$$

so that $\hat{\mathscr{H}}(\Sigma_1)$ is invariant under L_{f_1} and $L_{g_{1j}}$. But $\mathscr{H}(\Sigma_1)$ is defined as the smallest subspace of $C^\omega(M_1)$ with these properties. Hence $\hat{\mathscr{H}}(\Sigma_1) = \mathscr{H}(\Sigma_1)$.

For a fixed $\phi_1 \varepsilon \mathscr{H}(\Sigma_1)$, define now $\beta(\phi_1) = \{\phi_2 \varepsilon \mathscr{H}(\Sigma_2) ; \phi_2 = \phi_1$ on trajectories$\}$, and suppose that $\phi_2, \phi_2' \varepsilon \beta(\phi_1)$. Then ϕ_2 and ϕ_2' will agree on the reachable set $R(x_o)$ for some $x_o \varepsilon M_2$ which by hypothesis contains an open subset of M_2. By analytic continuation it follows that $\phi_2 = \phi_2'$ on M_2 and hence $\beta(\phi_1)$ is a singleton set. Thus β defines a linear map satisfying the theorem and is clearly unique.

b) Let $X \varepsilon \mathscr{L}(\Sigma_1)$ and define $\gamma(X)$ by

(1.2.2.) $$\beta(L_X(\psi)) = L_{\gamma(X)}(\beta(\psi)) \qquad \forall\, \psi \varepsilon \mathscr{H}(\Sigma_1).$$

Then, if $\gamma(X)$ exists it is clearly unique since, by observability,

$$L_Y(\phi) = d\phi(Y) = 0, \ \forall\phi\epsilon\mathscr{H}(\Sigma_2) \iff Y = 0$$

Linearity of γ follows from the linearity of β. We now define

$$\gamma(f_1) = f_2 \qquad \gamma(g_{1j}) = g_{2j} \qquad 1 \leq j \leq m$$

Then, by part a), these vector fields satisfy (1.2.2.). If (1.2.2.) is also true for some $X\epsilon\mathscr{L}(\Sigma_1)$ we see

$$\begin{aligned} \beta(L_{[f,X]}(\psi)) &= \beta(L_{f_1}(L_X\psi) - L_X(L_{f_1}\psi)) \\ &= L_{\gamma(f_1)}\beta(L_X\psi) - L_{\gamma(X)}\beta(L_{f_1}(\psi)) \\ &= L_{[\gamma(f_1),\gamma(X)]}(\beta(\psi)). \end{aligned} \tag{1.2.3.}$$

Thus (1.2.2.) is true for $[f_1,X]$ and similarly $[g_{1j},X]$. Since $\mathscr{L}(\Sigma_1)$ is generated by $\{f_1,g_{1j},\ldots,g_{1m}\}$ we see that γ extends to a linear map from $\mathscr{L}(\Sigma_1)$ on to $\mathscr{L}(\Sigma_2)$. The homomorphism property follows from (1.2.3.)

□

From this theorem we can deduce immediately that if Σ_1 and Σ_2 are both minimal then $\mathscr{L}(\Sigma_1)$ (resp.$\mathscr{S}(\Sigma_i)$, resp. $\mathscr{H}(\Sigma_i)$) is isomorphic to $\mathscr{L}(\Sigma_2)$ (resp.$\mathscr{S}(\Sigma_2)$, resp. $\mathscr{H}(\Sigma_2)$). Indeed, $\mathscr{H}(\Sigma_1)$ and $\mathscr{H}(\Sigma_2)$ are isomorphic if Σ_1 and Σ_2 are only accessible, although in this case observability of Σ_1 need not imply the observability of Σ_2 (consider, for instance, the two systems

$$\Sigma_1 \begin{cases} \dot{x}_1 = A_1x_1 + ub_1 \\ \dot{x}_2 = A_2x_2 + ub_2 \\ y = Cx_1 \end{cases} \qquad \& \qquad \Sigma_2 \begin{cases} \dot{x}_1 = A_1x_1 + ub_1 \\ \\ y = Cx_1 \end{cases}$$

both of which may be controllable, yet, for non-trivial A_2,b_2,Σ_1 cannot be observable).

Theorems (1.2.4) and (1.2.6) clearly show the importance of the objects $\mathscr{H},\mathscr{L},\mathscr{S}$ in systems theory – in particular with reference to questions of controllability, observability and realisability. However, these concepts are expressed in coordinate free terms and, whilst this condition certainly validates their use, in any given situation coordinates must be used for their calculation. Consequently, the question of the existence

$$L_Y(\phi) = d\phi(Y) = 0, \ \forall \phi \in \mathscr{H}(\Sigma_2) \iff Y = 0$$

Linearity of γ follows from the linearity of β. We now define

$$\gamma(f_1) = f_2 \qquad \gamma(g_{1j}) = g_{2j} \qquad 1 \leq j \leq m$$

Then, by part a), these vector fields satisfy (1.2.2.). If (1.2.2.) is also true for some $X \in \mathscr{L}(\Sigma_1)$ we see

$$\begin{aligned} \beta(L_{[f,X]}(\psi)) &= \beta(L_{f_1}(L_X\psi) - L_X(L_{f_1}\psi)) \\ &= L_{\gamma(f_1)}\beta(L_X\psi) - L_{\gamma(X)}\beta(L_{f_1}(\psi)) \\ &= L_{[\gamma(f_1),\gamma(X)]}(\beta(\psi)). \end{aligned} \tag{1.2.3.}$$

Thus (1.2.2.) is true for $[f_1,X]$ and similarly $[g_{1j},X]$. Since $\mathscr{L}(\Sigma_1)$ is generated by $\{f_1,g_{1j},..,g_{1m}\}$ we see that γ extends to a linear map from $\mathscr{L}(\Sigma_1)$ on to $\mathscr{L}(\Sigma_2)$. The homomorphism property follows from (1.2.3.)

$\square$

From this theorem we can deduce immediately that if Σ_1 and Σ_2 are both minimal then $\mathscr{L}(\Sigma_1)$ (resp. $\mathscr{S}(\Sigma_1)$, resp. $\mathscr{H}(\Sigma_1)$) is isomorphic to $\mathscr{L}(\Sigma_2)$ (resp. $\mathscr{S}(\Sigma_2)$, resp. $\mathscr{H}(\Sigma_2)$). Indeed, $\mathscr{H}(\Sigma_1)$ and $\mathscr{H}(\Sigma_2)$ are isomorphic if Σ_1 and Σ_2 are only accessible, although in this case observability of Σ_1 need not imply the observability of Σ_2 (consider, for instance, the two systems

$$\Sigma_1 \begin{cases} \dot{x}_1 = A_1x_1 + ub_1 \\ \dot{x}_2 = A_2x_2 + ub_2 \\ y = Cx_1 \end{cases} \quad \& \quad \Sigma_2 \begin{cases} \dot{x}_1 = A_1x_1 + ub_1 \\ \\ y = Cx_1 \end{cases}$$

both of which may be controllable, yet, for non-trivial A_2, b_2, Σ_1 cannot be observable).

Theorems (1.2.4) and (1.2.6) clearly show the importance of the objects $\mathscr{H}, \mathscr{L}, \mathscr{S}$ in systems theory - in particular with reference to questions of controllability, observability and realisability. However, these concepts are expressed in coordinate free terms and, whilst this condition certainly validates their use, in any given situation coordinates must be used for their calculation. Consequently, the question of the existence of some canonical set of coordinates arises - the basic philosophy being to choose the coordinates in such a way as to make the subsequent system representation as simple as possible in some (arbitrary, subjective) sense. Linear systems have many advantages from this point of view since any coordinate change is also linear. Hence, the full power of matrix algebra can be used resulting in, for example, the well-known controllable, or observable, companion forms. Similar descriptions have also been obtained in the nonlinear case although this is usually at the expense of some observability criterion mimicing further properties of linear systems (see for instance, Gauthier and Bornard [1], or Nijmeier [1] and the final chapter of this thesis).

At a more naive level, however, a natural question to ask would be whether it is possible to choose a chart so that, locally, the system algebra or observation space contains the relevant coordinate vector fields or coordinate functions. Such systems are said to be coordinate canonical. In respect to the algebra the following lemma is crucial (see also Jacubzyk and Respondek [1], in which a similar result is obtained for systems with no control term).

LEMMA (1.27)

Let Σ be a linear analytic system defined on a manifold M^n. Then if $x_o \epsilon M, \exists$ a chart (U,ϕ) about x_o s.t. on U, $\mathscr{L}^{\phi} \supset \{\frac{\partial}{\partial\phi_i}\ ;\ 1 \leq i \leq k\}$ iff $\mathscr{L}$ contains an abelian subalgebra, L_o, s.t. $\dim L_o(x) = k\ \forall\ x$ in some neighbourhood U' of x_o.

[$\mathscr{L}^{\phi}$ is the description of $\mathscr{L}$ in terms of the chart (U,ϕ)].

Proof

The implication (<=) is a direct consequence of $\text{Th}^{\text{m}}14$, chapter 5 in Spivak [1]. Conversely, since $\mathscr{L}$ is invariant under coordinate transformations and $\mathscr{L}_o^{\phi} = \text{Sp}\ \{\frac{\partial}{\partial\phi_i}\ ;\ 1 \leq i \leq k\}$ is abelian and spans a k-dimensional distribution the conclusion follows.

□

As examples of this lemma, consider the linear system on $\mathbb{R}^n$

$$(1.2.4) \qquad \dot{x} = Ax + \Sigma u_i b_i$$

Then $\mathscr{L} = \text{Sp}\{Ax, A^k b_i \;;\; 1 \leq i \leq m,\ 1 \leq k \leq n-1\}$. Thus, we can certainly choose a coordinate basis so that $\mathscr{L}$ contains $\{\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_\ell}\}$ where $\text{Sp}\{A^k b_i\} = \mathbb{R}^\ell$. In particular, if (1.2.4) is controllable, then $\mathscr{L}$ will be (controllably) coordinate canonical. However, the lemma also allows for the construction of counter examples showing that a system may not be coordinate canonical - for instance the system

$$\begin{cases} \dot{x}_1 = 1 \\ \dot{x}_2 = u_1 \\ \dot{x}_3 = x_2 + u_2 \end{cases}$$

has $\mathscr{L} = \text{Sp}\{\frac{\partial}{\partial x_1} + x_2 \frac{\partial}{\partial x_3}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\}$ which is non-abelian, consequently even under a nonlinear coordinate change, $\mathscr{L}$ cannot contain $\{\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}$ and $\frac{\partial}{\partial x_3}\}$.

In the light of these comments, the final result of this section (obtained independently by Fliess [1]) is quite remarkable.

THEOREM 1.2.8

Let Σ be an accessible, weakly observable linear analytic system on a connected manifold M. Then, $\forall x_o \in M$, there exists a coordinate description on a neighbourhood U_{x_o} of x_o such that the localised observation space contains the coordinate functions (ie Σ is observably coordinate canonical).

Proof

Since the system is assumed to be weakly observable, there exist $\omega_1, \ldots, \omega_n \in \mathscr{H}$ satisfying $\text{Sp}\{d\omega_1(x_o), \ldots, d\omega_n(x_o)\} = T^*_{x_o}M$. By the inverse function theorem, the map $\Omega: M \to \mathbb{R}^n$ defined by

$$\Omega_i(x) = \omega_i(x) \qquad 1 \leq i \leq n$$

is therefore an analytic diffeomorphism of a neighbourhood U of x_o onto an open set in $\mathbb{R}^n$. For a trajectory $x^u(t;t_o,x_o)$ of Σ, if we then define $z^u(t) = \Omega(x^u(t))$ it is readily seen that

$$\Sigma_z \quad \begin{cases} \dot{z}^u = (\Omega_* f)(z^u) + \Sigma u_i (\Omega_* g_i)(z^u) \\ y_j = h_j(\Omega^{-1}(z^u)) \end{cases}$$

is an accessible, weakly observable system defined on $\Omega(U)$, with the same input-output properties as Σ restricted to U. By Th^m (1.2.6a), $\forall \phi \varepsilon \mathscr{H}(\Sigma)$ $\exists ! \beta(\phi) \varepsilon \mathscr{H}(\Sigma_z)$ satisfying

$$\begin{aligned} \phi(\Omega^{-1}(z^u(t))) &= \phi(x^u(t)) \\ &= \beta(\phi)(z^u(t)). \end{aligned}$$

From which it follows that $\beta(\omega_i)(z^u(t)) = z_i^u(t)$ ie the functions giving the coordinates of a point on a trajectory of Σ_z are in $\mathscr{H}(\Sigma_z)$. Thus $z \to \beta(\Omega)(z)$ is the identity along trajectories in $\Omega(U)$. But Σ_z is accessible, so $\beta(\Omega)$ is the identity on some open subset of $\Omega(U)$. Hence, by analytic continuation, it is the identity on the whole of $\Omega(U)$, proving the claim.

□

§1.3 Graded Polynomial Systems

Having introduced much of the nomenclature of nonlinear systems theory and motivated the study of associated algebraic objects we now turn our attention to a novel class of systems which have hitherto not been studied for their own sake. These systems exhibit some extremely interesting and appealing global structure, yet at the same time are expressed in fairly simple almost canonical terms.

DEFINITION 1.3.1.

A linear analytic system

$$\Sigma \quad \begin{cases} \dot{x} = f(x) + \sum_{i=1}^{m} u_i g_i(x) \\ y_j = h_j \qquad\qquad 1 \leq j \leq p \end{cases}$$

defined on a manifold M is said to be in a graded polynomial form (g.p.f.) if

a) $M = \mathbb{R}^n$ and $\mathbb{R}^n$ is graded of degree N

b) W.r.t. this gradation $f \varepsilon V_0, g_i \varepsilon V_1$ and $h_j \varepsilon Q^{r_j}$, $1 \leq i \leq m$, $1 \leq j \leq p$

c) $f, g_1, \ldots, g_m$ are complete (this actually follows from the form of these vector fields given below).

$\square$

As a consequence of a further result of Palais', it follows that any vector field $X \in \mathscr{L}(\Sigma)$ is complete if Σ is in g.p.f. since $\mathscr{L} \subset V_o$ and V_o is finite dimensional. Using the decomposition of V_o and V_1 given in Th^m (1.1.2.), we see that any system in g.p.f. takes the form (for suitable polynomial vectors)

$$(1.3.1.) \quad \begin{cases} \dot{\underline{x}}_1 = A_1\underline{x}_1 + p_1 + \Sigma u_i g_{1i} & \underline{x}_i \in \mathbb{R}^{n_i} \quad 1 \leq i \leq N \\ \dot{\underline{x}}_2 = A_2\underline{x}_2 + p_2(\underline{x}_1) + \Sigma u_i g_{2i}(\underline{x}_1) & \\ \cdot & \\ \cdot & \\ \dot{\underline{x}}_N = A_N\underline{x}_N + p_N(\underline{x}_1, \ldots, \underline{x}_{N-1}) + \Sigma u_i g_{Ni}(\underline{x}_1, \ldots, \underline{x}_{N-1}) & \end{cases}$$

$$y_j = h_j(\underline{x}_1, \ldots, \underline{x}_N) \qquad \in Q^{r_j}$$

so that, if $N = 1$, any system with linear dynamics and polynomial output is in g.p.f. Once again, though we stress that if $\mathbb{R}^n = \bigoplus_{i=1}^{N} \mathbb{R}^{n_i}$, any n_i may be zero.

The structure theory of graded spaces developed in section 1.1. imposes certain restrictions on systems in g.p.f. We summarise some of the more straight forward ones here.

THEOREM 1.3.1.

Let Σ be in g.p.f. Then

(i) $\mathscr{S}$ is finite dimensional, nilpotent and of codimension ≤ 1 in $\mathscr{L}$.

(ii) $\mathscr{L}$ is solvable

(iii) There is a descending chain of subspaces $\{\tilde{H}^k;\ 0 \leq k \leq R+1\}$ of $\mathscr{H}$, satisfying

a) $\mathscr{H} = \tilde{H}^o \supset \tilde{H}^1 \supset \ldots \supset \tilde{H}^{R+1} = \{0\}$. b) $L_f(\tilde{H}^k) \subset \tilde{H}^k$

c) $L_{g_i}(\tilde{H}^k) \subset \tilde{H}^{k+1} \quad 1 \leq i \leq m$.

Moreover, if Σ is strongly accessible then $\exists$ q s.t. $\tilde{H}^q = \mathbb{R}$

(iv) If Σ is minimal (strongly accessible and weakly observable), then in the expansion (1.3.1.) we have

a) $\underline{x}_N \neq 0 \Rightarrow r_j \geq N$ for some $j \in \{1,..,p\}$

b) $\underline{x}_1 \neq 0 \Rightarrow \{g_{1i},\ 1 \leq i \leq m_1\} \neq \{0\}$.

Proof

(i) By definition, $\mathscr{S} = \{ad_f^k g_i;\ 1 \leq i \leq m,\ k \geq 0\}_{L.A.}$ so by previous comments $\mathscr{S}$ must be finite dimensional and of codim ≤ 1 in $\mathscr{L}$. Moreover, if $X \in V_1$, since $f \in V_0$ it follows that $[f,X] \in V_1$. Consequently, $\mathscr{S} \subset V_1$. But V_1 is nilpotent from the filtration properties of the sequence $\{V_j;\ j \leq p\}$. Hence, $\mathscr{S}$ is nilpotent.

(ii) $\mathscr{S}$ is a solvable ideal of $\mathscr{L}$ and $\mathscr{L}/_{\mathscr{S}}$ is abelian. Hence $\mathscr{L}$ is solvable.

(iii) Let $R = \max\limits_j \min\limits_{r_j} \{r_j; h_j \in Q^{r_j}\}$. Then, clearly, $\mathscr{H} \subset Q^R$ and $\tilde{H}^k \triangleq \mathscr{H} \cap Q^{R-k}$ satisfies (a,b,c). Let $q = \min\{k; \tilde{H}^{k+1} = \{0\}\}$ so $\tilde{H}^q \neq \{0\} = H^{q+i}$. But then $L_X(\phi) = 0\ \forall \phi \in \overset{k}{H^q}$ and $X \in \mathscr{S}$. This implies ϕ is constant along trajectories of $\mathscr{S}$ and hence, by s.a., ϕ is constant on an open subset of $\mathbb{R}^n$. By analytic continuation, ϕ is then constant on $\mathbb{R}^n$ and at least one element of $\tilde{H}^q$ is non zero by assumption. So $\tilde{H}^q = \mathbb{R}$.

(iv) a) Assume that $r_j < N$, $\forall 1 \leq j \leq p$. Then, if R is as defined in part (iii) we see that $R < N$. So

$$\begin{aligned} d\mathscr{H}_x \subset dQ_x^R &\subset \bigoplus_{j=1}^{R} (Q^{R-j} \oplus \Delta^j)_x \\ &\subset \bigoplus_{j=1}^{R} T_x^* \mathbb{R}^{n_j} \\ &\subsetneq T_x^* \mathbb{R}^n \end{aligned}$$

contradicting weak observability.

b) Similarly, if $g_{1i} = 0\ \forall i$, then

$$g_i \in \bigoplus_{j=2}^{N} Q^{j-1} \oplus \Delta_j \quad \text{and } f \in \bigoplus_{k=1}^{N} Q^k \oplus \Delta_k$$

Thus, $\mathcal{S} \subset [\bigoplus_{j=2}^{N} Q^{j-1} \otimes \Delta_j, \bigoplus_{k=2}^{N} Q^{k} \otimes \Delta_k] + [\bigoplus_{j=2}^{N} Q^{j-1} \otimes \Delta_j, Q^{1} \otimes \Delta_1]$

$$\subset \bigoplus_{j=2}^{N} Q^{j-1} \otimes \Delta_j$$

Consequently,

$$\mathcal{S}_x \subset \bigoplus_{j=2}^{N} T_x \mathbb{R}^{n_j}$$

contradicting strong accessibility.

$\square$

In the remainder of this section we study some of the deeper aspects of the algebraic nature of this class of systems. To begin with, recall that, as we have pointed out already, the class of minimal linear systems is both coordinate canonical and in g.p.f. Obviously, it is too much to expect that any g.p.f. is also coordinate canonical. However, by suitably enlarging $\mathcal{S}(\Sigma)$ and $\mathcal{H}(\Sigma)$ we obtain system invariants which do contain the relevant vector fields and functions. We defer discussion of the observability aspects of this question until the next chapter, and concentrate here on the strong accessibility algebra and system diffeomorphisms.

We start by considering the following example, defined on $\mathbb{R}^3 = \mathbb{R}^1 \oplus \mathbb{R}^1 \oplus \mathbb{R}^1$

$$(1.3.2.) \qquad \begin{cases} \dot{x}_1 = u \\ \dot{x}_2 = x_1^2 \\ \dot{x}_3 = x_2 + x_1^3 + x_2 u \end{cases} \qquad x(0) = 0$$

For which it is readily calculated that

$$\mathcal{S} = \mathrm{Sp}\,\{\frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}, \frac{\partial}{\partial x_1} + x_2\frac{\partial}{\partial x_3}, x_1\frac{\partial}{\partial x_2} + x_1^2\frac{\partial}{\partial x_3}, x_1\frac{\partial}{\partial x_3}\}$$

so, in particular $\frac{\partial}{\partial x_1} \notin \mathcal{S}$. However, if we generate $\mathcal{S}$ as a module over $\mathbb{R}[x_1,..,x_n]$ instead of as a real Lie algebra, then the coordinate vector fields will be elements of this new object, since $x_2\frac{\partial}{\partial x_3}$ will also be an element. In fact, for this example, we readily see that $\mathbb{R}[x_1,...,x_n] \otimes \mathcal{S} = D_1(\mathbb{R}^n)$ so (1.3.1.) is in this sense "algebraically" coordinate canonical. We now show that the structure exhibited by this example is also valid in general.

THEOREM 1.3.2.

Let Σ be a polynomial, minimal system in g.p.f. on $\mathbb{R}^n$. Then the Lie algebra $\mathbb{R}[x_1,..,x_n] \otimes \mathscr{S} = D_1(\mathbb{R}^n)$.

Proof

Let the state space decomposition be given by $\mathbb{R}^n = \bigoplus_{i=1}^{N} \mathbb{R}^{n_i}$ and define

$$K^i = \mathscr{S} + \{(P_{N-1} \otimes \Delta_N) \oplus \dots \oplus (P_{i-1} \otimes \Delta_i)\} \quad 1 \leq i \leq N$$

where $P_j = \mathbb{R}[\underline{x}_1,\dots,\underline{x}_j]$, and

$$K^{N+1} = \mathscr{S}.$$

Then, for $\phi \epsilon P_j$ and $X \epsilon P_{k-1} \otimes \Delta_k$ we see that $L_X(\phi) \epsilon P_j$ if $k \leq j$ and is identically zero for $j < k$. Then

$$[P_{K-1} \otimes \Delta_k, P_{\ell-1} \otimes \Delta_\ell] \subset P_m \otimes \Delta_m, \; m = \min\{k,\ell\}$$

from which we deduce that $(P_{N-1} \otimes \Delta_N) \oplus \dots \oplus (P_{i-1} \otimes \Delta_i)$ is a Lie algebra. Moreover, by Th^m (1.3.1.) $\mathscr{S} \subset V_1$, and clearly $V_1 \subset (P_{N-1} \otimes \Delta_N) \oplus ..\oplus (P_o \otimes \Delta_o)$.

Hence

$$\begin{aligned}[K^i,K^i] &\subset [\mathscr{S},\mathscr{S}] + [\mathscr{S},\overline{K}^i] + [\overline{K}^i,\overline{K}^i] \\ &\subset \mathscr{S}^2 + [V_1,\overline{K}^i] + (\overline{K}^i)^2\end{aligned}$$

where $\overline{K}^i = (P_{N-1} \otimes \Delta_N) \oplus .. \oplus (P_{i-1} \otimes \Delta_i)$, showing that K^i (and consequently $P_j \otimes K^i$) is a Lie algebra.

Assume, for the moment, that

$$\Delta_i \subset P_{i-1} \otimes K^{i+1} \qquad 1 \leq i \leq N \tag{1.3.3.}$$

Then, in particular, $\Delta_N \subset P_{N-1} \otimes K^{N+1} = P_{N-1} \otimes \mathscr{S}$. Thus,

$$\begin{aligned}P_{N-1} \otimes \mathscr{S} \subset P_{N-1} \otimes K^N &= P_{N-1} \otimes \mathscr{S} + P_{N-1} \otimes \{P_{N-1} \otimes \Delta_N\} \\ &\subset P_{N-1} \otimes \mathscr{S}.\end{aligned}$$

Similarly,

$$\begin{aligned} P_{N-1} \oplus K^r &= P_{N-1} \oplus \{K^{r+1} + P_{r-1} \oplus \Delta_r\} \\ &\subseteq P_{N-1} \oplus K^{r+1} + P_{N-1} \oplus \Delta_r \\ &\subseteq P_{N-1} \oplus K^{r+1} \qquad \text{by (1.3.3.)} \end{aligned}$$

We therefore see that $P_{N-1} \oplus K^1 = P_{N-1} \oplus \mathscr{S}$ and hence that $\Delta_1 \oplus .. \oplus \Delta_N \subset P_{N-1} \oplus \mathscr{S}$, from which the theorem follows. It remains to establish the validity of the identity (1.3.3.). First, note that by strong accessibility $\mathscr{S}$ is transitive. Further, $\mathscr{S}^2 \triangleq [\mathscr{S},\mathscr{S}] \subset V_2(\mathbb{R}^n)$ and can, therefore, only span $T_x\mathbb{R}^{n_2} \oplus ... \oplus T_x\mathbb{R}^{n_N}$. Thus, $\mathscr{S} \backslash \mathscr{S}^2$ must contain sufficient vector fields to span Δ_1. Since $\mathscr{S} \subset V_1$, we conclude that

$$(1.3.4.) \quad \mathscr{S} \supset \{\frac{\partial}{x_i} + \sum_{j=2}^{N} \sum_{k_j=1}^{n_j} r_{ij}^{k_j} (\underline{x}_1,..,\underline{x}_{j-1}) \frac{\partial}{\partial x^j_{k_j}} \; ; \; 1 \leq i \leq n_i\} .$$

with $r_j^{k_j} \epsilon Q^{j-1} \subset P_{j-1}$. Since $\mathscr{S} \subset K^2$, it follows that (1.3.3.) is true for $i = 1$. Let $N_r = n_1 + .. + n_r$. Then $\exists \; \hat{X}_1,...,\hat{X}_{N_r} \epsilon \mathscr{S}$ spanning $T_o\mathbb{R}^n$. From the polynomial nature of $\mathscr{S}$, we see that by possibly taking a suitable linear combination, these vector fields can take the form

$$\hat{X}_i = \frac{\partial}{\partial x_i} + \sum_{j=2}^{N} \sum_{k_j=1}^{n_j} r_{ij}^{k_j} (\underline{x}_1,..,\underline{x}_{j-i}) \frac{\partial}{\partial x^j_{k_j}} \qquad 1 \leq i \leq N_r$$

with $r_{ij}^{k_j} (0) = 0$, moreover, from the definition of K^{r+1}, we see that

$$(1.3.5.) \quad X_i = \frac{\partial}{\partial x_i} + \sum_{j=2}^{r} \sum_{k_j=1}^{n_j} r_{ij}^{k_j} \frac{\partial}{\partial x^j_{k_j}} \quad \epsilon \; K^{r+1}$$

Now assume that (1.3.3.) is valid for $1 \leq i < r$. It is readily seen that

$$K^{i+1} = K^{r+1} + \bigoplus_{j=i}^{r-1} P_j \oplus \Delta_{j+1}$$

from which we find (using the inductive hypothesis and noticing that $P_i \subset P_{i+1}$ and $P_j \oplus P_k = P_{\max\{k,j\}}$)

(1.3.6.) $\Delta_i \subset P_{i-1} \otimes K^{r+1} + \bigoplus_{j=i}^{r-1} P_j \otimes \Delta_{j+1}$

Since $\Delta_i \cap \bigoplus_{j=i}^{r-1} P_i \otimes \Delta_{i+1} = \{0\}$, it follows that (for $1 \leq i < r$)

(1.3.7.) $$\hat{Z}_{i,k} = \frac{\partial}{\partial x^i_k} + \sum_{j=i+1}^{r} \sum_{k_j=1}^{n_j} \psi^i_{k,k_j} \frac{\partial}{\partial x^j_{k_j}}$$

are elements of $P_{i-i} \otimes K^{r+1} \subset P_{r-1} \otimes K^{r+1}$, $\forall \frac{\partial}{\partial x^i_k} \in \Delta_i$ and some $\psi^i_{k,kj} \in P_{j-1}$. In turn, this implies that

(1.3.8.) $$Z_{i,k} = \frac{\partial}{\partial x^i_k} + \sum_{k_r=1}^{n_r} \hat{\psi}^i_{k,k_r} \frac{\partial}{\partial x^r_{k_r}} \qquad 1 \leq i \leq r-1$$

is an element of $P_{r-1} \otimes K^{r+1}$ for some $\hat{\psi}^i_{k,k_r} \in P_{r-1}$. For $\frac{\partial}{\partial x^{r-1}_k} \in \Delta_{r-1}$, (1.3.8.) is a trivial corollary of (1.3.7). To prove the general case, suppose it is true for $i = \ell,\dots,r-1$. Then, in particular, $\phi Z_k \in (P_j \otimes \Delta_i) + (P_{r-1} \otimes \Delta_r) \forall \phi \in P_j$, $j \leq r-1$. But the coefficients $\psi^{\ell-1}_{k,k_j}$ defining $\hat{Z}_{\ell-1,k}$ are polynomials in P_{j-1} for $j \leq r$. Consequently,

$$Z_{\ell-1,k} = \hat{Z}_{\ell-1,k} - \Sigma\Sigma\, \psi^{\ell-1}_{k,k_j} Z_{j,k_j}$$

is of the desired form, so (1.3.8.) is also valid for $i = \ell-1$. Similarly, by taking a suitable combination of X_i, as defined in (1.3.5), with $Z_{\ell,k}$, we see that

$$X_{j,k} = \frac{\partial}{\partial x^j_k} + \sum_{k_r=1}^{n_r} q^{k_r}_{j,k} \frac{\partial}{\partial x^r_{k_r}} \in P_{r-1} \otimes K^{r+1}, \quad 1 \leq j \leq r$$

From the above construction, it is also clear that $q^{k_r}_{j,k}(0) = 0$. Since $q^{k_r}_{j,k} \in P_{r-1}$ we can therefore, expand it as a polynomial in a single variable x^m_ℓ, with coefficients in $\mathbb{R}[x^1_1,\dots,\hat{x}^m_\ell,\dots,x^{n_{r-1}}_{r-1}]$ ($\hat{.}$ denoting exclusion) so

$$\begin{cases} q^{k_r}_{j,k} = \sum_{p=0}^{s} (x^m_\ell)^p \tilde{q}^{k_r}_{p,j,k} \\ \dfrac{\partial \tilde{q}}{\partial x_j} = 0 \end{cases}$$

But then, it is easily seen that

$$[X_{m,\ell}, X_{r,k}] = \left[\frac{\partial}{\partial x^m_\ell}, \sum_{k_r=1}^{n^r} \sum_{p=0}^{s} (x^m_\ell)^p \tilde{q}^{k_r}_{p,j,k} \frac{\partial}{\partial x^r_{k_r}}\right]$$

$$= \Sigma p \; (x^m_\ell)^{P-1} \tilde{q}^{k_r} \frac{\partial}{\partial x^r_{k_r}} .$$

We have already seen that $P_{r-1} \oplus K^{r+1}$ is a Lie algebra so this element is also in $P_{r-1} \oplus K^{r+1}$. Inductively, we see that

$$ad^s_{X_{m,}} (X_{r,k}) \Rightarrow \Sigma \tilde{q}^{k_r} \frac{\partial}{\partial x^r_{k_r}} \in P_{r-1} \oplus K^{r+1}$$

and, consequently, that $\frac{\partial}{\partial x^r_k} \in P_{r-1} \oplus K^{r+1}$. The induction on (1.3.3.) is therefore complete and the theorem is proved.

□

<u>Remark</u>: This theorem can be restated as " The modules over $\mathbb{R}[x_1, \dots, x_n]$ generated by $\mathscr{S}$ and $\{\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\}$ are identical". From standard results on module bases (Jacobson [1]) it follows that $\exists X_1, \dots, X_n \in \mathscr{S}$ and polynomials p_{ij} such that

$$\sum_{j=1}^{n} p_{ij} X_j = \frac{\partial}{\partial x_i} \tag{1.3.9.}$$

Let $P = [p_{ij}]$ and $V(x) = Sp\{X_1(x), \dots, X_n(x)\}$. Then, from (1.3.9.) we see that

$$P(x)\; V(x) \cong \mathbb{R}^n \qquad \forall\, x \in \mathbb{R}^n .$$

Consequently, both P(x) and V(x) must have full rank at all points $x \in \mathbb{R}^n$. From this, it follows that det $P(x) \neq 0$, so that $P(x)$ is invertible $\forall x$. However, X_j is also a polynomial combination of $\{\frac{\partial}{\partial x_i} \; ; \; 1 \leq i \leq n\}$ so

$P(x)^{-1}$ is also a polynomial matrix. This implies that $P(x)$ is, in fact, unimodular, ie det $P(x) =$ non-zero constant. We can therefore prove the following corollary, showing how far $\mathscr{S}$ is from being coordinate canonical.

COROLLARY 1.3.3.

Let Σ be as in Th^m(1.3.2) Then, if for some polynomial matrix P and corresponding module basis $\{X_1,..,X_n\}$ as described in (1.3.9.) there is a map Φ: $\mathbb{R}^n \to \mathbb{R}^n$ satisfying $D\Phi_x = P(x)$, $\mathscr{S}$ is coordinate canonical (globally).

Proof

Clearly, Φ must be polynomial and hence satisfies

$$||\Phi(x)|| \to \infty \quad \text{as} \quad ||x|| \to \infty$$

$$D\Phi_x \in GL(n;\mathbb{R}) \qquad \forall\, x\in\mathbb{R}^n.$$

These are precisely the conditions for Palais Global Inverse Function Theorem to apply. Hence, Φ is a diffeomorphism and we can therefore use it to define a further strongly accessible realisation, $\Phi_*\Sigma$, of Σ on $\mathbb{R}^n$. Moreover, Φ induces an isomorphism Φ_*: $\mathscr{S}(\Sigma)\to\mathscr{S}(\Phi_*\Sigma)$. Thus, $\forall$ $Y\in\mathscr{S}(\Sigma)$ using §1.2 we see

$$(\Phi_*Y) = \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} p_{ij} \, (\Phi^{-1}(z))Y_j(\Phi^{-1}(z)) \frac{\partial}{\partial z_i}$$

But, by assumption, $\exists$ $X_j\in\mathscr{S}(\Sigma)$, of the form $X_j = \Sigma X_{jk} \frac{\partial}{\partial x_k}$ and satisfying

$$\sum_{j=1}^{n} p_{ij}(x)X_{jk}(x) = \delta_{ki} \qquad \forall\, x\in\mathbb{R}^n.$$

From which we see that $\Phi_*X_i = \frac{\partial}{\partial z_i}$ and thus $\mathscr{S}$ is coordinate canonical as required. □

As a partial converse of this result, suppose that $\mathscr{S}$ is coordinate canonical. Then $\exists$ $X_1,..,X_n\in\mathscr{S}$, which commute and span $T_x\mathbb{R}^n$, $\forall x$. So, for a fixed $x_o\in\mathbb{R}^n$

$$\Sigma p_j(x_o)X_j(x_o) = 0 \iff p_j(x_o) = 0$$

If p_j are analytic functions, it therefore follows that X_j are independent over $C^\omega(\mathbb{R}^n)$ and hence form a module basis for both $D_1(\mathbb{R}^n)$ and $\Gamma^\omega(T\mathbb{R}^n)$. In particular, (1.3.9) is satisfied. However, this in itself is not enough to guarantee the existence of a suitable diffeomorphism Φ such that $D\Phi_x = P(x)$. For this to be true, it is equivalent to ask that the one forms defined by $\sum_{j=1} p_{ij}(x)dx^j$ be exact which, in turn, is equivalent to requiring that these forms are closed, ie they satisfy

$$\frac{\partial p_{ij}}{\partial x_k} = \frac{\partial p_{ik}}{\partial x_j} \qquad 1 \leq j,k \leq n.$$

As the following example shows, even if $X_1,..,X_n$ commute, $P(x)$ need not be exact: Let $X_1 = \frac{\partial}{\partial x_1} + x_1\frac{\partial}{\partial x_2}$, $X_2 = \frac{\partial}{\partial x_2}$. Then

$$P(x) = \begin{bmatrix} 1 & -x_1 \\ 0 & 1 \end{bmatrix}$$

so $\frac{\partial p_{11}}{\partial x_2} = 0 \neq \frac{\partial p_{12}}{\partial x_1} = -1$.

A further consequence of Theorem (1.3.2) is given in the following result

COROLLARY 1.3.4

Let Σ be as in Th^m(1.3.2) and define, for $k \geq 0$

$$\hat{D}_k(\mathbb{R}^n) = \{\sum_{|\alpha|\leq k} \phi_\alpha(x) \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}} ;\ \phi_\alpha \in \mathbb{R}[x_1,..,x_n], \alpha\in\mathbb{N}^n\}$$

$$A_k = A_k(\Sigma) = \sum_{j=0}^{k} \mathbb{R}[x_1,..x_n] \otimes \mathcal{S}^{\otimes j}$$

where $\mathcal{S}^{\otimes j} = \mathcal{S}\otimes \dots \otimes\mathcal{S}$, j-factors. Then $A_k = \hat{D}_k$, $\forall\ k \geq 0$.

Proof

By definition

$$\hat{D}_0\ (\mathbb{R}^n) = \mathbb{R}[x_1,..,x_n] = A_0$$

so the result is true for $k = 0$. For $k = 1$, the claim is also valid by Th^m(1.3.2). Inductively, therefore, assume it is true for $k = 0,...,N$.

Then, by the Leibnitz formula we see, for $\underline{x} = (x_1, \ldots, x_n)$

$$\begin{aligned} D_1 \otimes A_k &= \mathbb{R}[\underline{x}] \otimes \mathscr{S} \otimes A_k \\ &\subset \sum_{j=0}^{k} \mathbb{R}[\underline{x}] \otimes \mathscr{S}^{\otimes j+i} + \sum_{j=0}^{k} \mathbb{R}[\underline{x}] \otimes \mathscr{S}^{\otimes j} \\ &\subseteq A_{k+1} \end{aligned} \tag{1.3.10}$$

But, from the inductive hypothesis and Th^m (1.3.2) we have

$$\left\{\frac{\partial^k}{\partial x_{i_1}, \ldots, \partial x_{i_k}} ; 1 \leqslant i_j \leqslant n, 1 \leqslant j \leqslant k\right\} \subset A_k \qquad 0 \leqslant k \leqslant N \tag{1.3.11}$$

and $$\left\{\frac{\partial}{\partial x_i} ; 1 \leqslant i \leqslant n\right\} \subset A_1$$

so by (1.3.10), (1.3.11) is also true for $k = N + 1$. Since A_{N+1} is closed under multiplication by elements in $\mathbb{R}[\underline{x}]$, the result follows

$\square$

This result concludes our discussion of the strong accessibility algebra in this chapter. We now turn our attention to the structure of those diffeomorphisms on $\mathbb{R}^n$ preserving g.p.f's. As a preliminary observation, we prove the following lemma generalising the classical result stated at the end of §1.1, namely that for a vector field X and a smooth function ϕ

$$L_X(\phi) = 0 \iff \phi \text{ is constant along trajectories of } X. \tag{1.3.12}$$

LEMMA 1.3.5

Let $\mathbb{R}^n = \bigoplus_{i=1}^{N} \mathbb{R}^{n_i}$ and suppose w.r.t. this gradation $\mathscr{S}$ is a transitive Lie subalgebra of $V_1(\mathbb{R}^n)$. Then, if $\phi \in C^\omega(\mathbb{R}^n) \exists p = p(\phi)$ s.t. $\forall X_1, \ldots, X_p \in \mathscr{S}$ $L_{X_1} \ldots L_{X_p}(\phi) = 0 \iff \phi$ is polynomial.

Proof

($\Leftarrow$) is trivial since by definition $\phi \in Q^j$ for some j. The properties of $V_1(\mathbb{R}^n)$ then ensure the claim.

($\Rightarrow$) First, we note that if $\mathscr{L}_k$ is the space of vector fields of order $\leqslant k$ as defined in §1.1, then $V_k \subset \mathscr{L}_k \; \forall k \leqslant N$ and since $[\mathscr{S}, \mathscr{S}] \subset [V_1, V_1] \subset V_2$, it inductively follows that $\mathscr{S}^i \triangleq [\mathscr{S}^{i-1}, \mathscr{S}] \subset V_i$ and $S^{N+1} = \{0\}$. Hence

$X\varepsilon\mathscr{S} \Rightarrow X\varepsilon\mathscr{L}_k$ for some $1 \leq k \leq N$. Moreover, from the results of Sussman [2] and the transitivity of $\mathscr{S}$, the integral curves of the vector fields in $\mathscr{S}$ fill an open subset of $\mathbb{R}^n$. Thus if $p(\phi) = 1$, so $L_X(\phi) = 0 \ \forall X\varepsilon\mathscr{S}$, then, by (1.3.12), ϕ is constant on some open set in $\mathbb{R}^n$. By analytic continuation, it then follows that ϕ is constant everywhere.

Assume, then, that the result is true for $p = 1,\dots,K-1$ and let $\phi\varepsilon C^{\omega}(\mathbb{R}^n)$ satisfying

(1.3.13) $\quad L_{X_1}\dots L_{X_K}(\phi) = 0 \quad \forall X_i\varepsilon\mathscr{S},\ 1 \leq i \leq K.$

From the graded form of Taylors series we can decompose ϕ as

$$\phi = \phi_o + \hat{\phi}$$

with $\phi_o \ \varepsilon\ Q^{KN}$ and $\hat{\phi}\varepsilon C_{KN+1}$. (1.3.13) then becomes

$$0 = L_{X_1}..L_{X_K}(\phi_o) + L_{X_1}\dots L_{X_K}(\hat{\phi})$$

However, each $X_i\varepsilon V_{k_i} \subset \mathscr{L}_{k_i}$ with $1 \leq k_i \leq N$ so $\sum_{i=1}^{K} k_i \triangleq m \leq KN$. From the properties of V_{k_i} and $\mathscr{L}_{k_i}$ we see

(1.3.14) $\quad L_{X_1}..L_{X_K}(\phi_o)\ \varepsilon\ Q^{KN-m}$ and $L_{X_1}\dots L_{X_K}(\hat{\phi})\ \varepsilon\ C_{KN+1-m}$ so that

$KN+1-m > 0$. Since $Q^j \cap C_\ell = \{0\}$ for $0 \leq j \leq \ell$, it follows that (1.3.13) can hold iff both terms in (1.3.14) vanish. From (1.3.12) and transitivity of $\mathscr{S}$, we infer that $L_{X_2}\dots L_{X_K}(\hat{\phi})$ is constant $\forall X_2,\dots X_K$. However, in turn, this function is an element of $C_{\hat{m}}$, $\hat{m} = (KN + 1 - \sum_{i=2}^{K} k_i) > 0$ and hence must also be identically zero. From the inductive hypothesis, it follows that $\hat{\phi}$ is also polynomial and hence the theorem is proved.

$\square$

From Theorem (1.2.5) we know that if Σ_1 and Σ_2 are both minimal, linear analytic realisations of the same input-output map defined on manifolds M^m and N^n then $n = m$ and there is a unique analytic diffeormorphism $\Phi : M \rightarrow N$ 'preserving' trajectories and inducing an isomorphism between all the system algebras. As a corollary of the above lemma, we see that if Σ_1 and Σ_2 are both in g.p.f. then Φ is actually polynomial.

THEOREM 1.3.6.

Let Σ_1 and Σ_2 be minimal realisations of the same input-output map in g.p.f. and let $\Phi: \mathbb{R}^n \to \mathbb{R}^n$ be the system isomorphism described above. Then Φ is polynomial.

Proof

We let x_i, ∇_i denote the state variable and dynamics of Σ_i, $i = 1,2$. Then by definition, $\forall$ input u

$$\Phi(x_1^u(t)) = x_2^u(t)$$

Consequently, the systems

$$\Sigma_1' \begin{cases} \nabla_1 \\ \hat{y}_j(t) = \Phi_j(x_1) \end{cases} \quad \& \quad \Sigma_2' \begin{cases} \nabla_2 \\ \hat{y}_j(t) = x_2^j(t) \end{cases}$$

are also both minimal realisations of the same input-output map. Since the output of Σ_2' is polynomial (in fact linear) and by assumption $\mathscr{S}(\Sigma_2')$ is transitive, it follows from the above lemma that $\exists$ p s.t.

$$L_{X_1} \cdots L_{X_p}(x_2^j) = 0 \qquad \forall X_i \in \mathscr{S}(\Sigma_2') \quad 1 \leq i \leq p$$

From Theorem (1.2.6) we therefore see

$$L_{Y_1} \cdots L_{Y_p}(\Phi_j) = 0 \qquad \forall Y_i \in \mathscr{S}(\Sigma_1') \quad 1 \leq i \leq p.$$

But $\mathscr{S}(\Sigma_1')$ is also transitive so, by lemma (1.3.5) Φ_j is polynomial.

□

From the statement of the theorem it is trivial that Φ^{-1} is also polynomial (since this is a diffeomorphism between Σ_2 and Σ_1). This imposes immediate restrictions on Φ as it is well-known that not every polynomial function has a polynomial inverse - indeed, in the scalar case this is equivalent to requiring Φ to be affine-linear. The problem of classifying those diffeomorphisms of $\mathbb{R}^n$ satisfying these conditions and also carrying one graded structure onto another remains open.

CHAPTER II: CANONICAL REPRESENTATIONS OF G.P.F'S

In this chapter we continue the analysis of the g.p.f. by constructing two specific canonical representations of a minimal system in g.p.f. In doing so, we also establish a dual to Th^{m} (1.3.2), thus showing that any minimal g.p.f. is algebraically observable. Additionally, we show that any s.a. g.p.f. has an input-output map described by a stationary finite Volterra series. Conversely, a theorem of Crouch [1], shows that any s.f.v.s. has a minimal realisation in g.p.f. These remarks have two interesting corollaries. Firstly, they show that the class of s.a. systems in g.p.f. is closed under the operation of finding a minimal realisation. This contrasts markedly with other classes of nonlinearity - for instance in the bilinear case. Secondly, and of more importance, one may think of a s.f.v.s. as a sort of truncated Taylors series expansion of a smooth input-output map, or, indeed, as a generalised functional polynomial. As in the finite dimensional case, these polynomials have strong approximation properties - in fact, it has been shown that the set of s.f.v.s. is dense in the class of causal, stationary continuous input-output maps defined on a finite time interval and bounded controls, (Fliess [2], Sussmann [3]). Thus, over some suitable domain the class of systems in g.p.f. may be used to approximate to an arbitrary degree of accuracy any nonlinear system depending continuously on the input. The implications of this point, for identification are obvious.

This chapter is divided into two sections. In the first we construct a global coordinate chart for a minimal system in g.p.f. using the descending chain decomposition of the state space. The subsequent representation of the system is readily seen to be in observable coordinate canonical form. In particular, this implies that the algebra generated by the observation space is equal to the whole ring of polynomial functions, and it is then shown that as a consequence of Th^{m} (1.3.6), this is the case for any minimal g.p.f.

In the second section, the connection with the s.f.v.s. is explored. Indeed, an algebraic characterisation of s.a. realisations of finite Volterra series is given from which the above remarks follow immediately.

§2.1: The Graded Observable Polynomial Form

In the previous chapter we showed that minimal systems in g.p.f. need not be controllably coordinate canonical. To contrast this result, we prove here that the same class of systems is observably coordinate canonical; indeed we also show, as a dual to Th^{m} (1.3.2), that any such representation is algebraically o.c.c. in the sense that the algebra generated by the observation space is the ring of polynomial functions on $\mathbb{R}^{n}$. The approach taken is based on the descending chain of subspaces $\{\tilde{H}^{k};\ 0 \leq k \leq R\}$ of the observation space $\mathscr{H}$, from which we choose a global set of coordinates. The subsequent realisation will be seen to satisfy these requirements. We begin by proving a result, also of independent interest, essentially due to Crouch [1], and based on a technique of Hermann & Krener [1].

THEOREM 2.1.1

Let Σ be a strongly accessible linear analytic complete system $\{f, g_i, h_j;\ 1 \leq i \leq m,\ 1 \leq j \leq p\}$ on an analytic connected manifold M satisfying either

(i) $\mathscr{S}$ is nilpotent, with descending chain $\mathscr{S}^{\ell}$ $(\mathscr{S}^{1} = \mathscr{S}, \mathscr{S}^{\ell+1} \triangleq [\mathscr{S}, \mathscr{S}^{\ell}])$

or (ii) $\mathscr{H}$ has a descending chain $\mathscr{H} = \bar{H}^{0} \supset \bar{H}^{1} \supset \ldots \supset \bar{H}^{p+1} = \{0\}$ with

$$L_f(\bar{H}^{k}) \subset \bar{H}^{k} \text{ and } L_{g_j}(\bar{H}^{k}) \subset \bar{H}^{k+1} \qquad 0 \leq k \leq p+1$$

Then the distributions $x \to \mathscr{S}^{\ell}(x)$ and $x \to d\bar{H}^{k}(x)$ are constant dimensional. In particular, s.a. and w.o. are determined by evaluation at a single point.

Proof

Let S be the subgroup of Diff (M) generated by the trajectories of the vector fields in $\mathscr{S}$, so that

$$S = \{\gamma_1(t_1)\ldots.\gamma_n(t_n);\ \gamma_i(\cdot) \text{ is a trajectory of } X_i \epsilon \mathscr{S},\ t_i \epsilon \mathbb{R},\ 1 \leq i \leq n\}.$$

By strong accessibility and analyticity, S acts transitively on M, ie, $\forall$ $x_0, x_1 \epsilon M \ \exists \ \gamma \epsilon S$ s.t. $\gamma(x_o) = x_1$.

Now, for $X, Y \epsilon \Gamma^{\omega}(TM)$ and $\phi \epsilon C^{\omega}(M)$ the Campbell-Baker-Hausdorff formulae state that

$$(2.1.1) \qquad \gamma_X(-t)_* Y(\gamma_X(t)x_o) = \sum_{m \geq 0} \frac{t^m}{m!} L_X^m(Y)(x_o)$$

$$\gamma_X(-t)^* d\phi(\gamma_X(t)x_o) = \sum_{m \geq 0} \frac{t^m}{m!} dL_X^m(\phi)(x_o)$$

Hence, it follows that $\forall \gamma \epsilon \mathscr{S}$, dim $\mathscr{S}^{\ell}(\gamma(x_o) \leq \dim \mathscr{S}^{\ell}(x_o)$ (since $L_X(\mathscr{S}^{\ell}) \subset \mathscr{S}^{\ell+1} \forall X \epsilon \mathscr{S})$ and dim $d\bar{H}^k(\gamma(x_o)) \leq \dim d\bar{H}^k(x_o)$ (again $L_X(\bar{H}^k) \subset \bar{H}^{k+1} \forall X \epsilon \mathscr{S})$ by linearity of γ_* and γ^*. Since γ, x_o are arbitrary, the reverse inequalities are also valid and the theorem is proved.

$\square$

Remarks

(i) Note that neither $\mathscr{S}$ nor $\mathscr{H}$ themselves are required to be finite dimensional

(ii) If Σ is actually minimal we can associate with it two sets of indices, $\hat{n}_i = \dim \left(\mathscr{S}(x_o^i) \big/ \mathscr{S}^{i+1}(x_o) \right)$ and $\bar{m}_k = \dim \left(d\bar{H}^k(x_o) \big/ d\bar{H}^{k+1}(x_o) \right)$ which are invariant under diffeomorphism of the state space. Moreover, $\forall$ i or k $\exists$ $Y_1^q, \ldots, Y_{\hat{n}_i}^i \epsilon \mathscr{S}^i$ spanning $\mathscr{S}^i(x)$ and $\phi_1^1, \ldots, \phi_{m_k}^k \epsilon \bar{H}^k$ spanning $d\bar{H}^k(x)$, $\forall$ $x \epsilon M$. For if $\mathscr{S}$ has length q (so that $\mathscr{S}^{q+1} = \{0\}$), then by the C.B.H. formula (2.1.1) if $Y_1^q, \ldots, Y_{\hat{n}_q}^q$ span $\mathscr{S}^q(x_o)$ they must also span at every point in M since $L_X(Y_i^q) = 0 \ \forall \ X \epsilon \mathscr{S}$. Inductively, using the same argument, it is not difficult to see that a spanning set $\{Y_i^j\}$, with $1 \leq i \leq \hat{n}_j$, $q \geq j \geq \ell$, for

$\mathscr{S}^{\ell+1}(x_o)$ can be completed to a global spanning set for $\mathscr{S}^{\ell}(x_o)$ by vector fields $Y_1, \dots, Y^{\ell}_{\hat{n}_\ell}$ in $\dfrac{\mathscr{S}^{\ell}}{\mathscr{S}^{\ell+1}}$, and that these vector fields must form a basis for $\dfrac{\mathscr{S}^{\ell}(x)}{\mathscr{S}^{\ell+1}(x)}$ $\forall$ $x \in M$. The construction is identical for the observation space.

(iii) The integers $\bar{m}^k$ will depend on the choice of sequence of subspaces since we have not proved any uniqueness. For instance, if we define the length of the chain to be $\bar{p}$ where $\bar{p} = \min\{r;\ \bar{H}^{r+1} = \{0\}\}$ then there is no reason why any other chain satisfying the same conditions should have the same length. There is, however, a natural way of constructing each subspace to alleviate some of these problems by first setting $\hat{H}^o = \mathscr{H}$ and subsequently, if $\hat{H}^k$ is defined, we let

$$\hat{H}^{k+1} = \{L_f^j\, L_{g_i}(\phi)\ ;\ 1 \leq i \leq m,\ j \geq 0,\ \ \phi \in \hat{H}^k\}.$$

This sequence has the minimality property that for an arbitrary sequence $\{\bar{H}^k;\ 0 \leq k \leq \bar{p}\}$ we must have $\hat{H}^k \subset \bar{H}^k$ $\forall$ $0 \leq k \leq \bar{p}$, and $\hat{p} \leq \bar{p}$. The proof of these claims are straightforward. By definition, we know that $\hat{H}^o = \bar{H}^o = \mathscr{H}$ so we assume that for $0 \leq k \leq J$, $\hat{H}^k \subset \bar{H}^k$. Then $\forall$ $\phi \in \hat{H}^J$, it follows that $\phi \in \bar{H}^J$ and so

$$L_f^k\, L_{g_i}(\phi) \subset \hat{H}^{J+1} \cap \bar{H}^{J+1}$$

But $\hat{H}^{J+1}$ is generated in this fashion and so $\hat{H}^{J+1} \subset \bar{H}^{J+1}$, completing the induction. In particular, this means that $\hat{H}^{\bar{p}} \subset \bar{H}^{\bar{p}}$ and so $\hat{p} \leq \bar{p}$. The associated set of $\hat{p}$ integers $\{\hat{m}^k\}$ will be referred to as the observability indices of the system.

(iv) As a final comment on this result, note that the proof of Th^m(1.3.1) (iii c) also applies here, so that $\mathscr{H} \supset \bar{H}^{\bar{p}} = \hat{H}^{\hat{p}} = \mathbb{R}$.

The integer invariants defined above appear to have connections with the uniform unobservable structure defined by Nijmeier [1], but the precise nature of this relationship has yet to be established.

As a particular example of a class of systems satisfying the theorem we have, of course, the g.p. forms, which, indeed, satisfy both conditions. Moreover, in this case, from the particular polynomial structures involved it is clear that with respect to the associated graded decomposition $\mathbb{R}^n = \bigoplus_{j=1}^{N} \mathbb{R}^{n_j}$, $\mathscr{S}^\ell \subset V_\ell$ and $\hat{H}^k \subset \tilde{H}^k \subset Q^{R-k}$, $1 \leq \ell \leq N$, $0 \leq k \leq R$. Here R is the integer defined in Th^m (1.3.1) and $\{\tilde{H}^k;\ k \geq 0\}$ is an arbitrary sequence of subspaces satisfying the descending chain conditions. $\hat{H}^k$ is the minimal such sequence defined above. Thus,

$$\mathscr{S}^\ell(x_o) \subseteq T_{x_o}\Big(\bigoplus_{j=\ell}^{N} \mathbb{R}^{n_j}\Big) \qquad 1 \leq \ell \leq N \tag{2.1.2}$$

$$d\bar{H}^k(x_o) \subseteq T^*_{x_o}\Big(\bigoplus_{j=1}^{\min(N,R-k)} \mathbb{R}^{n_j}\Big) \qquad 0 \leq k \leq R-1. \tag{2.1.3}$$

where $\bar{H}^k$ is either $\tilde{H}^k$ or $\hat{H}^k$. We shall therefore say a minimal system in g.p.f. is in graded controllable polynomial form (g.c.p.f.) if equality holds in (2.1.2) at $x_o = 0$. Similarly, such a system is said to be in almost graded observable polynomial form (almost g.o.p.f.) if equality holds in (2.1.3) with $\bar{H}^k = \tilde{H}^k$, and in g.o.p.f. if equality holds for $\bar{H}^k = \hat{H}^k$, again with $x_o = 0$ but $R-N \leq k \leq R-1$. From Th^m (2.1.1) it follows that in either of these cases, equality will hold then for all $x_o \in \mathbb{R}^n$. It is our intention to show that any minimal system in g.p.f. also has realisations in (almost) g.o.p.f. and g.c.p.f. although these representations need not coincide. The latter situation is treated in the next section, but before we consider the former, we prove a simple lemma.

<u>Lemma 2.1.2</u>

Suppose *D* is a k-dimensional codistribution on an n-dimensional manifold M and that $\psi_1,\ldots,\psi_n \in C^\infty(M)$ satisfy

$$\left.\begin{array}{ll}\text{(i)} & d\psi_1,\ldots,d\psi_k \text{ span } D(x) \\ \text{(ii)} & d\psi_1,\ldots,d\psi_n \text{ span } T^*_xM\end{array}\right\} \quad \forall x \in U,\ U \text{ open in } M$$

Let $\phi \varepsilon C^{\infty}(M)$ such that $d\phi(x) \varepsilon \mathcal{D}(x) \forall x \varepsilon U$. Then $\exists\ \tilde{\phi}: \mathbb{R}^k \to \mathbb{R}$ s.t. $\phi(x) = \tilde{\phi}(\psi_1(x), \ldots, \psi_k(x))$.

Proof

By the inverse fuction theorem it is immediate that (by possible restriction) $\Psi: M \to \mathbb{R}^n$ defined by $\Psi_i(x) = \psi_i(x)$ is a diffeomorphism of U onto an open set in $\mathbb{R}^n$. Consequently, $\exists\ \hat{\phi}: \mathbb{R}^n \to \mathbb{R}$ $(\hat{\phi} = \phi \circ \Psi^{-1})$ such that $\phi(x) = \hat{\phi} \circ \Psi(x)$. By the chain rule, we see then

$$d\phi = \sum_{i=1}^{k} \frac{\partial \hat{\phi}}{\partial x_i} d\psi_i + \sum_{i=k+1}^{n+1} \frac{\partial \hat{\phi}}{\partial x_i} d\psi_i$$

But $d\phi(x) \varepsilon \mathcal{D}(x)$ so that the second sum is identically zero on U by (i) Linear independence of $d\psi_{k+1} \ldots, d\psi_n$ then implies that

$$\frac{\partial \hat{\phi}}{\partial x_i} = 0 \qquad k+1 \leq i \leq n$$

and, hence, $\hat{\phi} = \hat{\phi}(\psi_1, \ldots, \psi_k)$ as required.

(Clearly, if all the data is analytic then U can be taken to be a connected component of M).

□

We are now in a position to prove the major theorem of this section namely the construction of an almost g.o.p.f. for any minimal g.p.f. We remark, however, that the preliminary stages of the proof also apply to any minimal system satisfying Th^{m}(2.1.1 (ii)), thus giving an (almost) canonical local description for such systems.

THEOREM 2.1.3

Let Σ be a minimal system in g.p.f. Then Σ also has a minimal realisation in almost g.o.p.f., $\Sigma_{a.o}$. Moreover, $\mathcal{H}(\Sigma_{a.o})$ will contain all the coordinate functions and, in particular, Σ will also have a realisation in g.o.p.f.

Proof

We adopt the same notation as in Th^{m}(1.3.1) so that for the given g.p.f. $\mathcal{H}$ is assumed to be a subset of Q^R and there is a decomposition

$\mathscr{H} = \tilde{H}^{0} \supset \tilde{H}^{1} \supset \dots \supset \tilde{H}^{q} = \mathbb{R} \supset \{0\}$ with respect to the decomposition of the state space $\mathbb{R}^{n} = \bigoplus_{j=1}^{N} \mathbb{R}^{n_j}$ so that $R \geq N$. Now choose $\phi_{1}^{q-1},\dots,\phi_{m_{q-1}}^{q-1} \in \tilde{H}^{q-1}$ to span $d\tilde{H}^{q-1}(0)$ and hence by the remarks following Th^{m} (2.1.1) spanning $d\tilde{H}^{q-1}(x)$ $\forall x \in \mathbb{R}^{n}$. This can be completed to a basis for $d\tilde{H}^{q-2}(0)$ by functions $\phi_{1}^{q-2},\dots,\phi_{m_{q-2}}^{q-2} \in \tilde{H}^{q-2}\backslash\tilde{H}^{q-1}$ and, inductively, to a basis $\phi_{1}^{q-1},\dots,\phi_{m_{q-M}}^{q-M}$ for $d\mathscr{H}(0)$, which again spans the whole of $d\mathscr{H}(x)$ $\forall x \in \mathbb{R}^{n}$. Since each ϕ_{i}^{j} is polynomial, it therefore follows that $\phi:\mathbb{R}^{n}\to\mathbb{R}^{n}$ defined by

$$\phi(x) = \begin{bmatrix} \phi_{1}^{q-1}(x) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \phi_{m_{q-M}}^{q-M}(x) \end{bmatrix}$$

satisfies Palais' G.I.F.T. and is hence an analytic diffeomorphism. (In the more general case Φ is, if course, only a smooth diffeomorphism on some open subset of the state space).

Now let $x^{u}(t)$ be a trajectory of the g.p.f. so that it satisfies

$$\dot{x}^{u}(t) = f(x^{u}(t)) + \sum_{j=1}^{m} u_j g_j(x^{u}(t))$$

$$y_i(t) = h_i(x^{u}(t)) \qquad 1 \leq i \leq p$$

Then

$$\frac{d}{dt}\,\phi_{j}^{k}(x^{u}(t)) = L_f(\phi_{j}^{k})(x^{u}(t)) + \Sigma u_j\, L_{g_j}(\phi_{j}^{k})(x(t))$$

But $\phi_{j}^{k} \in \tilde{H}^{k}$, and $L_f(\tilde{H}^{k}) \subset \tilde{H}^{k}$, $L_{g_i}(\tilde{H}^{k}) \subset \tilde{H}^{k+1}$. Moreover, the codistributions $x \to d\tilde{H}^{k}(x)$ satisfy the conditions of Lemma (2.1.2) Hence, $\exists$ analytic functions satisfying

$$L_f(\phi_{j}^{k})(x^{u}(t)) = F_{j}^{k}(\phi_{1}^{q-1}(x^{u}(t)),\dots,\phi_{m_k}^{k}(x^{u}(t)))$$

$$L_g(\phi_j^k)(x^u(t)) = G_j^k(\phi_1^{q-1}(x^u(t)),\ldots,\phi_{m_{k-1}}^{k-1}(x^u(t)))$$

So, by defining, for $1 \leq k \leq M$

$$\underline{z}_k^u(t) = (\phi_1^{q-k}(x^u(t)),\ldots,\phi_{m_{q-k}}^{q-k}(x^u(t)))^T$$

we obtain a minimal system on $\mathbb{R}^n$ of the form

$$\Sigma_z \left\{ \begin{array}{ll} \dot{\underline{z}}_1 = F^1(\underline{z}_1) + \sum_{j=1}^{m} u_j G_j^1 & \in \mathbb{R}^{m_{q-1}} \\ \dot{\underline{z}}_2 = F^2(\underline{z}_1,\underline{z}_2) + \Sigma u_j G^2(\underline{z}_p) & \\ \vdots \qquad \vdots \qquad \vdots & \\ \dot{\underline{z}}_M = F^M(\underline{z}_1,\ldots,\underline{z}_M) + \Sigma u_j G_j^M(\underline{z}_1,\ldots,\underline{z}_{M-1}) & \in \mathbb{R}^{m_{q-M}} \\ y_i = (h_i \circ \Phi^{-1})(z) = H_i(z) & 1 \leq i \leq p. \end{array} \right.$$

By following the proof of $Th^m(1.2.8)$ it is clear that $\mathcal{H}(\Sigma_z)$ must contain the coordinate functions (and, hence, all the components of the vector fields $F = (F^1,..,F^M)^T$ and $G_j = (G_j^1,..,G_j^M)^T$). Further, the isomorphism $\beta: \mathcal{H}(\Sigma) \to \mathcal{H}(\Sigma_z)$ clearly induces a descending chain $\{\tilde{\mathcal{O}}^k;\ 0 \leq k \leq q\}$ of subspaces of $\mathcal{H}(\Sigma_z)$ satisfying

$$L_F(\tilde{\mathcal{O}}^k) \subset \tilde{\mathcal{O}}^k,\ L_G(\tilde{\mathcal{O}}^k) \subset \tilde{\mathcal{O}}^{k+1}\ \&\ \beta(\tilde{H}^k) = \tilde{\mathcal{O}}^k,\ \tilde{\mathcal{O}}^q = \mathbb{R}.$$

From which we see that $z \to z_k^i \in \tilde{\mathcal{O}}^{q-k}$ and that $d\tilde{\mathcal{O}}^{q-k}(z) = T_z^*(\bigoplus_{j=1}^{M} \mathbb{R}^{m_{q-j}})$ as required. (Also note, $\phi \in \tilde{\mathcal{O}}^{q-k} \Rightarrow \phi = \phi(\underline{z}_1,..,\underline{z}_k)$)

The realisation Σ_z is locally valid for any linear analytic minimal system satisfying the descending chain condition of $Th^m(2.1.1)$. It remains to show that in this specific situation, with Σ in g.p.f, the data H_i, F, and G_j are in the relevant spaces w.r.t. the decomposition $\mathbb{R}^n = \bigoplus_{j=1}^{M} \mathbb{R}^{m_{q-j}}$.

From the previous comments it suffices to show that $\tilde{\mathcal{O}}^k \subset Q^{q-k}$, $q-M \leq k \leq q$ for then we will have proved that $F \in V_0$ & $G_j \in V_1$, implying that $\mathcal{S}(\Sigma_z) \subset V_1$. By strong accessibility, the descending chain condition and Lemma (1.3.5) it then readily follows that the output functions H_i are also polynomial.

We have already seen that $\tilde{O}^q = \mathbb{R} = Q^o$. Inductively, we then assume that $\tilde{O}^{q-j} \subseteq Q^j$ for $0 \leq j \leq k-1$. By strong accessibility it is immediate that we can find vector fields $X^i_j \varepsilon \mathscr{S}(\Sigma)$, $1 \leq i \leq \bar{m}_j$, $1 \leq j \leq k_j$ spanning $T_z(\mathbb{R}^{\bar{m}_1} \oplus \,..\, \oplus \mathbb{R}^{\bar{m}_k})$ $1 \leq k_j \leq k$ where $\bar{m}_k \triangleq m_{q-k}$, and satisfying

$$(2.1.4) \qquad X^i_j = \frac{\partial}{\partial z^i_j} + \sum_{\ell=1}^{k} \sum_{r=1}^{\bar{m}_i} \psi^{ij}_{r\ell} \frac{\partial}{\partial z^r} + \hat{X}^i_j$$

$$(2.1.5) \qquad L_{\hat{X}^i_j}(z^r_\ell) = 0 \qquad\qquad \forall\ \ell \leq k$$

$$(2.1.6) \qquad \psi^{ij}_{r\ell}\ (0) = 0.$$

Moreover, since the coordinates $z \rightarrow z^r_\ell$ are elements of $\tilde{O}^{q-\ell}$ it follows from the induction assumption and (2.1.5) that $\psi^{ij}_{r\ell} \,\varepsilon\, Q^{\ell-1}$. Thus

$$Y^i_j \triangleq X^i_j - \hat{X}^i_j \qquad \varepsilon\, V_1(\mathbb{R}^{\bar{m}_1} \oplus \,..\, \oplus \mathbb{R}^{\bar{m}_k}) \qquad ,$$

and $\mathscr{L}'_k = \{Y^i_j;\ 1 \leq i \leq \bar{m}_j,\ 1 \leq j \leq k\}_{L.A.}$ acts transitively on $\mathbb{R}^{\bar{m}_1 + \,..\, \bar{m}_k}$. Further, for $\ell \leq k$ (2.1.5) also implies $L_{Y^i_j}(\tilde{O}^{q-\ell}) \subset \tilde{O}^{q-\ell-1}$, so $\mathscr{L}'_k$ satisfies all the conditions of Lemma (1.3.5) with $\phi \varepsilon \tilde{O}^{q-k}$. Hence, $\tilde{O}^{q-k} \subset Q^K$ for some integer K, and $K \geq k$ since $z \rightarrow z^r_k \,\varepsilon\, \tilde{O}^{q-k} \cap Q^k$. Now, fix $\underline{z}_i = \underline{z}^o_i$ for $i \neq j$, $1 \leq i \leq k$ and if $\phi \varepsilon\ \tilde{O}^{q-k}$ define

$$\phi_o(z) = \phi(\underline{z}^o_1, \ldots, \underline{z}_j, \ldots, \underline{z}^o_k)$$

Then

$$L_{Y^i_j}(\phi_o) = \frac{\partial}{\partial z^i_j}(\phi_o) + \Sigma \psi^{ij}_{rj} \frac{\partial}{\partial z^r_j} \phi_o$$

and $\psi^{ij}_{rj} = \psi^{ij}_{rj}\ (z^o_1, \ldots, z^o_{j-1})$, so are constant. But, by construction $\{Y^i_j;\ 1 \leq i \leq \bar{m}_j\}$ span $T_z(\mathbb{R}^{\bar{m}_j})$ $\forall z \varepsilon \mathbb{R}^{\bar{m}_1} \oplus \,..\, \oplus \mathbb{R}^{\bar{m}_k}$. Thus $\exists$ constants $\alpha^\ell_i = \alpha^\ell_i(z^o_1, \ldots, z^o_{j-1})$ satisfying

$$\sum_{r=1}^{\bar{m}_j} \alpha^\ell_r\, L_{Y^r_j}(\phi_o) = \frac{\partial}{\partial z^\ell_j}(\phi_o)$$

This implies that $\frac{\partial\phi_o}{\partial z_j^{\ell}} \varepsilon\ \tilde{O}^{q-k+1}$ and, inductively that $\frac{\partial^{|L|}\phi_o}{\partial z_j^{\ell_1}\ldots\partial z_j^{\ell_\sigma}} \varepsilon\, \tilde{O}^{q+\sigma-k}$

But $\tilde{O}^{q-(j-1)} = \tilde{O}^{q-k+(k-j-1)} \subset Q^{j-1}$ and so

$$\frac{\partial^{|L|}\phi_o}{\partial z_j^{\ell_1}\ldots\partial z_j^{\ell_\sigma}} = 0 \qquad\qquad \forall |L| \geq k-j-1$$

Hence ϕ_o is polynomial of degree $\leq k-j$. Since this is true for arbitrary $\underline{z}_i^o$ and $\underline{z}_j$ it follows that $\phi \varepsilon Q^k$ and the induction is complete, thus proving the theorem.

□

Since Σ_z is observably coordinate canonical and in g.p.f., it is immediately obvious that $\mathcal{H}_A(\Sigma_z)$, the algebra generated by $\mathcal{H}(\Sigma_z)$ under the pointwise operations of multiplication, is equal to $\mathbb{R}[z_1^1,\ldots,z_M^{m_M}]$. It is our intention to prove that this is in fact the case for any minimal system in g.p.f. First though, we note that the isomorphism β between the observation spaces of two minimal analytic realisations $(\Sigma_1$ & $\Sigma_2)$ of the same system extends to an algebra isomorphism. The details are trivially verified: surjectivity is obvious, for if $\Phi \varepsilon \mathcal{H}_A(\Sigma_2)$ then $\Phi = \Sigma A_\alpha \phi_\alpha^{\alpha_1} \ldots \phi_n^{\alpha_n}$ for some $A_\alpha \varepsilon \mathbb{R}$ and $\phi_1,\ldots,\phi_n \varepsilon \mathcal{H}(\Sigma_2)$. Hence $\Sigma A_\alpha (\beta^{-1}(\phi_1))^{\alpha_1} \ldots (\beta^{-1}(\phi_n))^{\alpha_n} \varepsilon \mathcal{H}_A(\Sigma_1)$ is mapped onto Φ by the algebraic extension $\hat{\beta}$ of β. On the other hand, suppose $\Psi \varepsilon \mathcal{H}_A(\Sigma_1)$ and that $\hat{\beta}(\Psi) = 0$. Then since $\Psi = \Sigma G_\gamma \psi_1^{\gamma_1} \ldots \Psi_m^{\gamma_m}$, with $\psi_1,\ldots,\psi_m$ linearly independent, we have

$$\hat{\beta}(\Psi) = \Sigma G_\gamma \beta(\psi_1)^{\gamma_1} \ldots \beta(\psi_m)^{\gamma_m} = 0$$

But β is an isomorphism so each product $\beta(\psi_1)^{\gamma_1} \ldots \beta(\psi_m)^{\gamma_m}$ is also linearly independent. Thus, each $G_\gamma = 0$ and so $\psi = 0$, and $\hat{\beta}$ is injective.

From these remarks we immediately see that if Σ is a minimal g.p.f. on $\mathbb{R}^n$ then $\mathcal{H}_A(\Sigma)$ is isomorphic to $\mathbb{R}[z_1,\ldots,z_n]$ and, moreover, is contained in $\mathbb{R}[x_1,\ldots,x_n]$, where x is the state variable of Σ. In itself, this is not a sufficient condition to imply the equality claimed for we need only consider the case $n = 1$, and the diffeomorphism

$\gamma(x) = z = \frac{x^3}{3} + x$. Then the algebra generated by $(\frac{x^3}{3} + x)$ is isomorphic to the algebra generated by z but is strictly contained in $\mathbb{R}[x]$. However, from Theorem (1.3.6), we also know that β is induced by a polynomial diffeomorphism with a polynomial inverse ($\gamma(x) = \frac{x^3}{3} + x$ does not satisfy this property). The following result then proves our claim that $\mathcal{A}_A(\Sigma) = \mathbb{R}[x_1,..,x_n]$.

THEOREM 2.1.4

Let A be a subalgebra of $\mathbb{R}[x_1,...,x_n]$ and $\gamma:\mathbb{R}^n \to \mathbb{R}^n$ be a polynomial diffeomorphism such that $\beta_\gamma : A \to \mathbb{R}[z_1,...,z_n]$ defined by $\beta_\gamma(p)(z) = p(\gamma^{-1}(z))$ is an algebra isomorphism. Then $A = \mathbb{R}[x_1,...,x_n]$ iff γ has polynomial inverse.

Proof

($\Rightarrow$) If $A = \mathbb{R}[x_1,...,x_n]$ then $\exists\, q_i \in \mathbb{R}[z_1,..,z_n]$ such that

$$\beta_\gamma^{-1}(q_i)(x) \triangleq x_i \qquad 1 \leq i \leq n.$$

But $\beta^{-1} = \beta_{\gamma^{-1}}$; so $\beta_\gamma^{-1}(q_i)(x) = q_i(\gamma(x))$. Setting $Q = (q_1,..,q_n)^T$ we see then that $Q \circ \gamma(x) = x$. Further,

$$\beta_\gamma^{-1}(q_i)(Q(z)) = q_i(\gamma \circ Q(z)) = q_i(z) \qquad \forall\, z.$$

so $\gamma \circ Q = Id$. Hence $Q = \gamma^{-1}$ and so γ has polynomial inverse

($\Leftarrow$) Since γ has polynomial inverse $\exists\, q_i \in \mathbb{R}[z_1,..,z_n]$ such that

$$q_i(z) = x_i$$

But then $\beta_\gamma^{-1}(q_i)(x) = q_i(\gamma(x)) = x_i$ and $\beta_\gamma^{-1}(q_i) \in A$. Thus, A contains all the coordinate functions and so $A = \mathbb{R}[x_1,..,x_n]$.

□

§2.2. Finite Volterra Series and G.P.F's

In this section we complete the analysis begun in the previous section by showing that any minimal system in g.p.f. also has a realisation in g.c.p.f. In doing so, we also provide an algebraic characterisation of

$\gamma(x) = z = \frac{x^3}{3} + x$. Then the algebra generated by $(\frac{x^3}{3} + x)$ is isomorphic to the algebra generated by z but is strictly contained in $\mathbb{R}[x]$. However, from Theorem (1.3.6), we also know that β is induced by a polynomial diffeomorphism with a polynomial inverse ($\gamma(x) = \frac{x^3}{3} + x$ does not satisfy this property). The following result then proves our claim that $\mathscr{H}_A(\Sigma) = \mathbb{R}[x_1,..,x_n]$.

THEOREM 2.1.4

Let A be a subalgebra of $\mathbb{R}[x_1,...,x_n]$ and $\gamma:\mathbb{R}^n \to \mathbb{R}^n$ be a polynomial diffeomorphism such that $\beta_\gamma : A \to \mathbb{R}[z_1,...,z_n]$ defined by $\beta_\gamma(p)(z) = p(\gamma^{-1}(z))$ is an algebra isomorphism. Then $A = \mathbb{R}[x_1,...,x_n]$ iff γ has polynomial inverse.

Proof

($\Rightarrow$) If $A = \mathbb{R}[x_1,...,x_n]$ then $\exists\, q_i \in \mathbb{R}[z_1,..,z_n]$ such that

$$\beta_\gamma^{-1}(q_i)(x) \overset{\Delta}{=} x_i \qquad 1 \leq i \leq n.$$

But $\beta^{-1} = \beta_{\gamma^{-1}}$ so $\beta_\gamma^{-1}(q_i)(x) = q_i(\gamma(x))$. Setting $Q = (q_1,..,q_n)^T$ we see then that $Q \circ \gamma(x) = x$. Further,

$$\beta_\gamma^{-1}(q_i)(Q(z)) = q_i(\gamma \circ Q(z)) = q_i(z) \qquad \forall\, z.$$

so $\gamma \circ Q = Id$. Hence $Q = \gamma^{-1}$ and so γ has polynomial inverse

($\Leftarrow$) Since γ has polynomial inverse $\exists\, q_i \in \mathbb{R}[z_1,..,z_n]$ such that

$$q_i(z) = x_i$$

But then $\beta_\gamma^{-1}(q_i)(x) = q_i(\gamma(x)) = x_i$ and $\beta_\gamma^{-1}(q_i) \in A$. Thus, A contains all the coordinate functions and so $A = \mathbb{R}[x_1,..,x_n]$.

□

§2.2. Finite Volterra Series and G.P.F's

In this section we complete the analysis begun in the previous section by showing that any minimal system in g.p.f. also has a realisation in g.c.p.f. In doing so, we also provide an algebraic characterisation of minimal realisations of so-called stationary finite Volterra series (s.f.v.s.) thereby establishing that systems in g.p.f. must have a s.f.v.s.

We begin by recalling that a finite Volterra series of length q is a functional mapping taking the form

$$y(t) = \sum_{k=0}^{q} \hat{W}_k(u)(t)$$

where

$$\hat{W}_k(u)(t) = \int_0^t \int_0^{\sigma_1} \ldots \int_0^{\sigma_{k-1}} W_k(t,\sigma_1,\ldots,\sigma_k)(u(\sigma_1),\ldots,u(\sigma_k))d\sigma_k \ldots d\sigma_1$$

and $W_k(t,\sigma_1,\ldots,\sigma_k)$ is a multilinear map on $\mathbb{R}^m \times \ldots \times \mathbb{R}^m \rightarrow \mathbb{R}^p$ (thus each function u is $\mathbb{R}^m$ valued), depending analytically on $t,\sigma_1,\ldots,\sigma_k$ with components $W_k^{ji_1 \ldots i_k}$, $1 \leq j \leq p$, $1 \leq i_\ell \leq m$, $1 \leq \ell \leq k$. The series is said to be stationary if, in addition

$$\left(\frac{\partial}{\partial t} + \sum_{i=1}^{k} \frac{\partial}{\partial \sigma_i} \right) W_k(t,\sigma_1,\ldots,\sigma_k) \equiv 0 \qquad 1 \leq k \leq q$$

and

$$W_0(t) \equiv 0$$

It is readily seen that a complete linear analytic system of the form

$$(2.2.1) \quad \begin{cases} \dot{x} = f(x) + \sum_{i=1}^{m} u_i g_i(x) & x(0) = x_0 \quad x \in M \\ y_j = h_j(x) & 1 \leq j \leq p \end{cases}$$

has an input-output map described by a (possibily infinite) Volterra series, with convergence guaranteed for suitably bounded controls. Moreover, the kernels, W_k, of this series (which will obviously depend on x_0) are given inductively by

$$(2.2.2) \quad \begin{cases} W_0^j(t,x_0) = h_j(\gamma_f(t)x_0) \\ W_k^{ji_1 \ldots i_k}(t,\sigma_1 \ldots,\sigma_k;x_0) = L_{\gamma_f}(-\sigma_k)_* g_{i_k}(W_{k-1}(t,\sigma_1,\ldots,\sigma_{k-1};\cdot)(\gamma_f(\sigma_k)x_0) \end{cases}$$

(Krener & Lesiak [1], Crouch [1]). A realisation of the form (2.2.1) is stationary if it is autonomous and $f(x_0) = 0$. The converse problem of finding conditions under which a given finite Volterra series can be represented as the solution to a control system (2.2.1) is also well-understood. We begin by stating the following fundamental result, generalising a similar criteria for linear systems.

THEOREM (2.2.1) (Brockett [1])

A (stationary) f.v.s. has a realisation by a (stationary) system (2.2.1) on an analytic manifold M^n iff it has a (stationary) bilinear realisation on some Euclidean space. Further, such an input-output map is realisable iff the kernels, W_n, are (stationary and) differentiably separable, ie each component of W_k can be written as a finite sum of products taking the form $\gamma_1(t)\gamma_2(\sigma_1)\ldots\gamma_k(\sigma_k)$ and each function γ_i is analytic.

□

Unfortunately, the bilinear realisation guaranteed by the above theorem need not be minimal. Indeed, if we consider the stationary case, then we can assume that the system can be represented as

$$(2.2.3)\quad \Sigma_z \quad \begin{cases} \dot{z} = (A + \Sigma u_i B_i) & z(0) = 0 \\ y_i = C_i z \end{cases}$$

with A and B_i constant matrices. $\mathscr{S}(\Sigma_z)$ will then consist entirely of linear vector fields so that, in particular,

$$X \epsilon \mathscr{S}(\Sigma_z) \Rightarrow X(0) = 0 \Rightarrow \mathscr{S}(\Sigma_2)(0) = \{0\}.$$

Hence, (2.2.3) can never be strongly accessible. This leads naturally to the problem of classifying the minimal realisations of such input-output maps, a question which has been neatly answered by a theorem due to Crouch [1] establishing immediately a point of contact with the previous work of this thesis.

THEOREM (2.2.2) (Crouch [1])

A s.f.v.s. of length q which has a linear analytic realisation (2.2.1) with complete vector fields has a minimal realisation in g.c.p.f. Moreover, w.r.t. the induced gradation on the (Euclidean) state space, each output function $h_j \varepsilon Q^q$.

□

In particular, it follows from Th^m (1.2.6), that if (2.2.1) is strongly accessible then its observation space is isomorphic to that of a system in g.p.f. and hence must be finite dimensional and satisfy a descending chain condition. It turns out that these conditions are also sufficient for a linear analytic system to have a s.f.v.s.. The proof presented here was developed independently of a similar result of Fliess and Kupka [1] for bilinear systems.

THEOREM 2.2.3

A strongly accessible, complete linear analytic system of the form (2.2.1) has a s.f.v.s. of length q, $q \geq 1$ iff the following conditions are satisfied

(i) $\mathscr{H}$ is finite dimensional

(ii) $\mathscr{H}$ has a descending chain of subspaces $\{\theta^k;\ 0 \leq k \leq q+1\}$ with

a) $\mathscr{H} = \theta^0 \supset \theta^1 \supset .. \supset \theta^{q+1} = \{0\}$, $\theta^q = \mathbb{R}$

b) $L_f(\theta^k) \subset \theta^k$, $L_{g_i}(\theta^k) \subset \theta^{k+1}$ $\quad 1 \leq i \leq m$.

Proof

(=>) From the preceding remarks, we know that $\mathscr{H}$ satisfies both (i) and (ii) except that the length of the chain of subspaces may not equal the length of the Volterra series. However, iterated use of the Campbell-Baker-Hausforff formula and the inductive formulae for the kernels shows that, about $t = \sigma_1 = \sigma_2 = .. = \sigma_k = 0$, W_k can be expanded as

$$W_k^{ji_1..i_k}(t,\sigma_1,..,\sigma_k;x_o) = \sum_{|\alpha|\geq 0} \frac{\sigma_k^{\alpha_k}\ldots\sigma_1^{\alpha_1} t^{\alpha_o}}{\alpha_k!\ldots\alpha_o!} L_{ad_f^{\alpha_k}(g_{i_k})} \ldots . L_{ad_f^{\alpha_1}(g_{i_1})} L_f^{\alpha_o}(h_j)(x_o)$$

$$= \sum_{|\alpha|\geq 0} \phi_\alpha^k(x_o)\frac{\sigma_k^{\alpha_k} .. t^{\alpha_o}}{\alpha!},$$

$$\phi_\alpha^k = \phi_\alpha^k(j,i_1,\ldots,i_k) \qquad k \geq 1$$

with $\phi_\alpha^k \varepsilon \mathscr{H}$, and $ad_f^o g = g$, $ad_f^{k+1}(g) = [f, ad_f^k g]$. Now, define $\theta^k = \{\phi_\alpha^\ell(j, i_1 \ldots, i_\ell);\ 1 \leq j \leq p,\ 1 \leq i_r \leq N,\ 1 \leq r \leq \ell,\ \ell \geq k,\ |\alpha| \geq 0\}$, so $\hat{\theta}^k \supset \hat{\theta}^{k+1} \supset \ldots \supset \hat{\theta}^{q+1} = \{0\}$. Then, it is clear that $\mathbb{L}_{g_i}(\hat{\theta}^k) \subset \hat{\theta}^{k+1}$ and inductive use of the formula $L_f L_g = L_{[f,g]} + L_g L_f$ shows that $L_f(\hat{\theta}^k) \subset \hat{\theta}^k$. Further, $h_j \varepsilon \hat{\theta}^{o|}$, $1 \leq j \leq p$. Hence $\hat{\theta}^o = \mathscr{H}$ since $\hat{\theta}^{o|}$ is invariant under L_f and L_{g_i} and any $\phi \varepsilon \hat{\theta}^{o|}$ is, by definition, in $\mathscr{H}$. Moreover, it is shown in Crouch [1], that W_q is independent of x_o, so that ϕ_α are all constant functions, at least one of which is non-zero (otherwise $W_q \equiv 0$). Thus $\hat{\theta}^q = \mathbb{R}$ and the proof is complete.

(<=) We now suppose that Σ is a strongly accessible system satisfying (i) and (ii). To show that Σ has a s.f.v.s. we construct a stationary bilinear realisation of the same input-output map, which by Th^m (2.2.1) will establish the claim.

Since $\mathscr{H}$ is finite dimensional, we may construct a basis $\{\phi_o, \phi_1^{q-1}, \ldots, \phi_{m_o}^o\}$ by first choosing $\phi_o \varepsilon \theta^q \backslash \{0\}$. Then, if $\{\phi_o, \ldots, \phi_{m_k}^k\}$ has been selected to span θ^{q-k} it is completed to a basis for θ^{q-1-k} by $\phi_1^{k+1} \ldots \phi_{m_{k+1}}^{k+1} \varepsilon \theta^{q-1-k} \backslash \theta^{q-k}$. Next, we define

$$z_k(t) = (\phi_1^{q-k}(x(t)), \ldots, \phi_{m_{z-k}}^{q-k}(x(t)))^T \qquad 0 \leq k \leq q$$

where x(t) is a trajectory of (2.2.1). Then

$$\frac{d\phi_j^{q-k}}{dt}(x(t)) = L_f(\phi_j^{q-k})(x(t)) + \Sigma u_k L_{\ell}(\phi_j^{q-k})(x(t))$$

But, by assumption $L_f(\phi_j^{q-k}) \varepsilon \theta^{q-k}$ and $L_{g_\ell}(\phi^{q-k}) \varepsilon \Theta^{q+1-k}$ and hence can be written as linear (constant coefficient) combinations of the basis functions. We then see that

$$(2.2.6) \quad \begin{cases} \dot{z}_o = 0 = A_{oo} z_o \\ z_1 = A_{1o} z_o + A_{11} z_1 + \Sigma u_\ell b_{1\ell} z_o \\ \cdot \\ \cdot \\ \dot{z}_q = A_{qo} z_o + A_{q1} z_1 + \ldots + A_{qq} z_q + \Sigma u_\ell (b_{q\ell} z_o + \ldots + B_{q-1\ell} z_{q-1}) \end{cases}$$

which can clearly be written as a bilinear system

$$\dot{z} = (A + \Sigma u_\ell B_\ell)z$$

with A a block triangular matrix and each B_ℓ strictly lower block triangular. Moreover, for an initial condition $x(0)$ for (2.2.1) we can readily arrange that $z(0)$ satisfies $A(z(0)) = 0$ by a simple coordinate translation: Let $\hat{z}(0) = \Phi(x(0))$, Φ has components $(\phi_1^q,\ldots,\phi_{m_o}^o)$. Then $\phi_1^q \neq 0$ (since it must span $\theta^q = \mathbb{R}$) but $A_{\infty}(\phi_1^q) = 0$. So if we define $\tilde{z}_0(t) = z_0(t)$, $\tilde{z}_k(t) = z_k(t) - \hat{z}_k(0)$, $1 \leq k \leq q$ then $\tilde{z}_k(0) = 0$ and so $A\tilde{z}(0) = 0$, whilst $\tilde{z}_k$ still satisfies (2.2.6).

Finally, it is obvious that since $h_j \in \mathscr{H}$, the outputs of (2.2.1) can be written as linear combinations of $(z_o,\ldots,z_q)$ and so

$$\begin{aligned} y_j &= C_j z \\ \dot{z} &= (A + \Sigma u_\ell B_\ell)z \qquad z(0) = 0 \end{aligned}$$

is the required, stationary, bilinear realisation of (2.2.1). The finiteness of the associated Volterra series is guaranteed by the nilpotence of B_ℓ, $1 \leq \ell \leq m$ and the solvability of A (Brockett [1]). □

We state as a trivial corollary of this result the culmination of the analysis begun in the previous section, namely a dual to Th^m (2.1.3).

COROLLARY 2.2.4

Let Σ be a strongly accessible system in g.p.f. Then Σ has a minimal realisation in g.c.p.f.

Proof

From Th^{ms} (1.3.1) and (2.2.3), Σ has a stationary finite Volterra series. Thus by Th^{ms} (2.2.2) and Σ has a minimal realisation in g.c.p.f.

□

To summarise, we have now characterised the g.p.f. in terms of its input-output map and have shown that there exist two specific minimal realisations, also in g.p.f. From the remarks following Th^m (2.1.1), this implies that $\exists$ two sets of integers, in turn characterising the associated s.f.v.s., namely

$$\hat{m}_k = \dim \left. d\hat{H}^k(0) \middle/ d\hat{H}^{k+1}(0) \right. \qquad \hat{n}_i = \left. \mathscr{S}^i(0) \middle/ \mathscr{S}^{i+1}(0) \right.$$

Further, these indices are the dimensions of the components of the gradations of the state spaces in the g.o. and g.c. polynomial forms respectively.

If Σ is an arbitrary, minimal linear analytic realisation of the Volterra series, then the associated controllability indices, $\hat{n}_i$, can be calculated directly from the kernels using an algorithm developed by Crouch [ ], and generalising the well-known factorisation $W_1(t,\sigma) = H(t)G(\sigma)$ for linear systems. On the other hand, the observability index $\hat{m}_k$, is readily seen to depend only on the k^{th} kernel. Indeed, if $\{\hat{H}^k; 0 \leq k \leq q\}$ are defined as in Th^m (2.2.3) then $\hat{m}_k$ is determined by $\hat{\theta}^k / \hat{\theta}^{k+1}$ and this quotient contains only those observable functions which are coefficients in the Taylors series expansion of W_k.

In general, the two sets of integers $\{\hat{n}_k\}$ and $\{\hat{m}_k\}$ need not be identical, for if we consider the system

$$(2.2.7) \qquad \begin{cases} \dot{x}_1 = u \\ \dot{x}_2 = x_1^2 \\ \dot{x}_3 = x_2 \end{cases} \qquad x(0) = 0$$

$$y = x_3 + x_1 x_2$$

which is clearly in g.p.f. on $\mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2}$, $n_1 = 1$, $n_2 = 2$ then

$$f = x_1^2 \frac{\partial}{\partial x_2} + x_2 \frac{\partial}{\partial x_3}, \quad g = \frac{\partial}{\partial x_1}, \quad h(x) = x_3 + x_1 x_2$$

and it is readily calculated that

$$\mathscr{S} = \mathscr{S}^1 = \mathrm{Sp}\{ g, x_1 \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}, x_1 \frac{\partial}{\partial x_3} \}$$

$$\mathscr{S}^2 = \mathrm{Sp}\{ \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3} \} \qquad \mathscr{S}^3 = \{0\}.$$

Hence, (2.2.7) is in g.c.p.f.

On the other hand,

$$\mathscr{H} = \mathrm{Sp}\{x_3 + x_1x_2,\ x_2, x_1, x_1^2, x_1^3, 1\}.$$

with $\hat{\theta}^1 = \mathrm{Sp}\{x_2, x_1^2, \hat{\theta}^2\}$, $\hat{\theta}^2 = \mathrm{Sp}\{x_1, \hat{\theta}^3\}$, $\hat{\theta}^3 = \mathbb{R}$.

Thus $\hat{m}_1 = \hat{m}_2 = \hat{m}_3 = 1$, and (2.2.7) is not in g.o.p.f. The algorithm for constructing the g.o.p.f. implies that by setting

$$z_1(t) = x_1(t),\ z_2(t) = x_2(t) \qquad z_3(t) = x_3(t) + x_1(t)x_2(t)$$

and differentiating, we will obtain the required g.o.p.f. Performing these operations yields

$$\begin{cases} \dot{z}_1 = u \\ \dot{z}_2 = z_1^2 \\ \dot{z}_3 = \dot{x}_3 + \dot{x}_1x_2 + x_1\dot{x}_2 = x_2 + ux_2 + x_1^3 \end{cases}$$

$$\qquad = z_2 + z_1^3 + uz_2$$

$$y = z_3(t) \qquad\qquad z(0) = 0.$$

and it is readily seen that this sytem is not in g.c.p.f.

For Volterra series with only one kernel, W_q, the situation is quite different. Indeed, it is shown in Crouch [1] (§4) that in fact $\hat{m}_k = \hat{n}_k\ \forall\, 0 \leq k \leq q$. As a simple example consider the minimal system with a single kernel of degree 2

$$\begin{cases} \dot{x}_1 = u \\ \dot{x}_2 = x_1^2 \end{cases} \qquad\qquad x(0) = 0$$

$$y = x_2$$

Then $\mathscr{S}^2 = \mathrm{Sp}\{\frac{\partial}{\partial x_2}\}$, $\mathscr{S}^1 = \mathrm{Sp}\{\mathscr{S}^2, x_1 \cdot \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_1}\}$, and

$$\hat{\theta}^3 = \mathbb{R},\ \hat{\theta}^2 = \mathrm{Sp}\{\hat{\theta}^2, x_1\},\ \hat{\theta}^1 = \mathrm{Sp}\{\hat{\theta}^2, x_2, x_1^2\}$$

so $\hat{m}_1 = \hat{n}_1 = \hat{m}_2 = \hat{n}_2 = 1$ as required.

We conclude this section, and indeed the chapter, with some remarks on feedback in nonlinear systems, and introduce an algebraic structure which

Hence, (2.2.7) is in g.c.p.f.

On the other hand,

$$\mathscr{H} = \mathrm{Sp}\{x_3 + x_1x_2,\ x_2, x_1, x_1^2, x_1^3, 1\}.$$

with $\hat{\theta}^1 = \mathrm{Sp}\{x_2, x_1^2, \hat{\theta}^2\}$, $\hat{\theta}^2 = \mathrm{Sp}\{x_1, \hat{\theta}^3\}$, $\hat{\theta}^3 = \mathbb{R}$.

Thus $\hat{m}_1 = \hat{m}_2 = \hat{m}_3 = 1$, and (2.2.7) is not in g.o.p.f. The algorithm for constructing the g.o.p.f. implies that by setting

$$z_1(t) = x_1(t),\ z_2(t) = x_2(t) \qquad z_3(t) = x_3(t) + x_1(t)x_2(t)$$

and differentiating, we will obtain the required g.o.p.f. Performing these operations yields

$$\begin{cases} \dot{z}_1 = u \\ \dot{z}_2 = z_1^2 \\ \dot{z}_3 = \dot{x}_3 + \dot{x}_1x_2 + x_1\dot{x}_2 = x_2 + ux_2 + x_1^3 \end{cases}$$

$$= z_2 + z_1^3 + uz_2$$

$$y = z_3(t) \qquad\qquad z(0) = 0.$$

and it is readily seen that this sytem is not in g.c.p.f.

For Volterra series with only one kernel, W_q, the situation is quite different. Indeed, it is shown in Crouch [1] (§4) that in fact $\hat{m}_k = \hat{n}_k\ \forall 0 \leq k \leq q$. As a simple example consider the minimal system with a single kernel of degree 2

$$\begin{cases} \dot{x}_1 = u \\ \dot{x}_2 = x_1^2 \end{cases} \qquad x(0) = 0$$

$$y = x_2$$

Then $\mathscr{S}^2 = \mathrm{Sp}\{\frac{\partial}{\partial x_2}\}$, $\mathscr{S}^1 = \mathrm{Sp}\{\mathscr{S}^2, x_1\frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_1}\}$, and

$$\hat{\theta}^3 = \mathbb{R},\ \hat{\theta}^2 = \mathrm{Sp}\{\hat{\theta}^2, x_1\},\ \hat{\theta}^1 = \mathrm{Sp}\{\hat{\theta}^2, x_2, x_1^2\}$$

so $\hat{m}_1 = \hat{n}_1 = \hat{m}_2 = \hat{n}_2 = 1$ as required.

We conclude this section, and indeed the chapter, with some remarks on feedback in nonlinear systems, and introduce an algebraic structure which is invariant under polynomial output feedback. Specifically, if the system in question takes the form (2.2.1) we assume that the input can be written as

$$u(t) = \gamma(y_1(t),..,y_p(t)) + \bar{u}$$

where $\gamma : \mathbb{R}^p \to \mathbb{R}^m$ is an arbitrary polynomial function, and $\bar{u} \in \mathscr{U}$. We then find that the system takes the form

$$\Sigma_\gamma \quad \begin{cases} \dot{x}(t) = (f + \Sigma\gamma_i(h_1(x(t)),..,h_p(x(t)))g_i) + \Sigma\bar{u}_i g_i \\ y_j(t) = h_j(x(t)). \end{cases}$$

Thus, the linear-analytic structure is preserved but any finer detail, such as bilinearity or a graded polynomial form, may be destroyed. However, note that since γ_i is polynomial in $h_1 \ldots h_p$, $1 \leq i \leq m$, $\mathscr{X}(\Sigma_\gamma) \subset \mathscr{X}_A(\Sigma)$ and $\mathscr{L}(\Sigma_\gamma) \subset \mathscr{X}_A(\Sigma) \oplus \mathscr{L}(\Sigma)$. From these identities we see immediately that

$$\mathscr{X}_A(\Sigma_\gamma) \subset \mathscr{X}_A(\Sigma) \ \& \ \mathscr{X}_A(\Sigma_\gamma) \oplus \mathscr{L}(\Sigma_\gamma) \subset \mathscr{X}_A(\Sigma) \oplus \mathscr{L}(\Sigma).$$

By applying the 'inverse feedback' $\bar{u} = -\gamma(y(t)) + u(t)$ to Σ_γ, we conclude that the reverse conclusions are also valid and, hence, that both $\mathscr{X}_A$ and $\mathscr{X}_A \oplus \mathscr{L}$ are invariant under this class of operations. (For minimal systems in g.p.f. this is almost trivial since Th^{ms}(1.3.2), (2.1.3) and (2.1.4) show that in this case $\mathscr{X}_A \oplus \mathscr{L} = D_1(\mathbb{R}^n)$).

The effects of such polynomial feedback on nonlinear systems have yet to be fully understood although some interesting features are already emerging. For instance, conditions have been derived under which the resulting system Σ_γ has a linear input-output map, with γ a linear function, (Nijmeier[2], Cyrot-Nomand & Monaco [1]). In similar vein, it may be possible to choose an output feedback so that $\mathscr{X}(\Sigma_\gamma) = \mathscr{X}_A(\Sigma)$ or, indeed, so that Σ_γ is controllably coordinate canonical. As a simple example of this behaviour, consider the system

$$\Sigma_1 \quad \begin{cases} \begin{cases} \dot{x}_1 = u \\ \dot{x}_2 = x_1^2 \end{cases} \\ y = x_2^3 \end{cases}$$

Under the linear feedback $u = y + \bar{u}$, this system becomes

$$\Sigma_2 \quad \begin{cases} \begin{cases} \dot{x}_1 = x_2^3 + \bar{u} \\ \dot{x}_2 = x_1^2 \end{cases} \\ y = x_2^3 \end{cases}$$

For which $\mathscr{L}(\Sigma_2) \supset \{ \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, x_1 \frac{\partial}{\partial x_2}, x_2 \frac{\partial}{\partial x_1}, x_2^2 \frac{\partial}{\partial x_1} \}$. This example also exhibits a further possible type of behaviour namely that $\mathscr{H}(\Sigma_2) = \mathscr{H}_A(\Sigma_1)$, for Σ_1 is minimal and in g.p.f. w.r.t. the decomposition $\mathbb{R}^2 = \mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2}$ $n_1 = n_2 = 1$. Hence, $\mathscr{H}_A(\Sigma_1) = \mathbb{R}[x_1,x_2]$. But $\mathscr{H}(\Sigma_2)$ clearly contains $\{1, x_2, x_2^2, x_2^3, x_1, (x_1 x_2)\}$. If we assume inductively that $\mathscr{H}(\Sigma_2)$ contains $x_1^{n-p} x_2^p$ for $0 \leq p \leq n$, then $L_{x_2^2 \frac{\partial}{\partial x_1}} (x_1 x_2^{n-1}) = x_2^{n+1} \in \mathscr{H}(\Sigma_2)$ and iterated applications of $L_{x_1 \frac{\partial}{\partial x_2}}$ show that $x_1^{n+1-p} x_2^p \in \mathscr{H}(\Sigma_2)$ for $\forall\ 0 \leq p \leq n+1$. Hence $\mathscr{H}(\Sigma_2) = \mathscr{H}_A(\Sigma_1) = \mathbb{R}[x_1,x_2]$. The general validity of this behaviour remains to be established.

CHAPTER III: ALGEBRAIC ESTIMATION

If the success of an Applied Mathematical theory were to be measured in terms of physical or 'real world' applications, then one of Control Theory's outstanding contributions must be the Kalman-Bucy filtering algorithm. The appeal and applicability of this method seems to lie at two levels. Besides the obvious advantage of answering a previously difficult problem by presenting a relatively simple, and thus more readily implementable scheme (in context the major tool available before the developments of Kalman & Bucy was the rather cumbersome theory due to Wiener [1](see also Kailaths paper [1])), the Kalman filter also has a conceptually attractive interpretation as an (apocryphal) Black-Box into which one inputs the observational data and which outputs the desired (optimal) estimate. A major drawback of the algorithm is, of course, that its use is restricted to linear systems and it is natural to ask if there are more general versions available which can handle nonlinearities; the answer is, luckily, in the affirmative. Of the several possible alternatives, probably the most famous are the equations of motion of the moments of the relevant conditional density, or of the evolution of the density itself. These results have been available in the literature since the mid 1960's (Wonham [1], Kushner [1], Jazwinskii [1], Bucy & Joseph [1]) but have recently received a more rigorous, general treatment through Martingale analysis (Lipster & Shiryaev [1], Kallianpur [1]). These methods have not yet achieved the same degree of popularity as the Kalman filter partly due to the increased mathematical maturity required to understand them, but in large part this shortcoming can be ascribed to an inherent element of infinite dimensionality preventing ready assimilation in software terms. Thus, having asked the question can the Kalman filter be generalised, we must now ask if there are generalisations which can be used as practical schemes. This problem forms the basis for the next two chapters.

As a starting point we take an equation due variously to Zakai [1], Mortensen [1] and Duncan [1], describing the evolution of an unnormalised version of the conditional density. The appeal of this approach is that despite the infinite dimensional nature of this equation (as will be seen, c.f. equn. (3.1.6), it is a stochastic partial differential equation) it is both bilinear and recursive and, moreover, conditional statistics can be expressed, simply, in terms of its solution. Thus, we can retain the intuitive idea of a filter as some system transforming observations into estimates. The point remaining unanswered is if this scheme has any practical significance since the problem of dimensionality is still present. A direct approach towards a solution can be made, and significant advances have been made by Davis [1], [2], using the ideas of Doss and Sussmann on the pathwise solution of stochastic systems. Here, however, we take a different point of view, suggested originally by Brockett [2] but which has since generated considerable interest and research activity (see for instance the proceedings Hazewinkel & Willems [1]), and study only the algebraic complexity of the filter defined by Zakai's equation.

Some justification for this methodology is presented in the first section of this chapter and it will be seen that the basic idea is to regard any more computationally efficient scheme as a lower (finite) dimensional realisation of the input-output map generated by the above Zakai system. Heuristically, we can then argue that the results of §1.2, and in particular Th^{m} (1.2.6), should still apply. In this fashion we arrive at the fundamental question treated in algebraic estimation theory, namely when is there a Lie algebra homomorphism between a Lie algebra consisting of differential operators on $\mathbb{R}^n$, and a Lie algebra of vector fields on a finite dimensional manifold? These ideas have been placed in a rigourous context by Hijab [1] but it is not difficult to see, indeed we shall show, how they work in most of the cases where 'practical' algorithms for the filtering problem can be found. Our primary objective in this chapter is, however, to study only one aspect of the algebraic estimation problem and that is to ask when the Lie algebra Λ is finite dimensional. The most important consequence of this hypothesis is that the unnormalised conditional density equation can be 'exactly modelled' by a bilinear stochastic differential system, but unfortunately it also imposes severe restrictions on the system generating the data. In the final section, therefore, as well as including several examples specialising the necessary conditions derived in §3.1, we also make some comments on the case that Λ is infinite dimensional although we defer our major excursion into this realm until the final chapter.

§3.1 Finite Dimensional Estimation Algebras

As mentioned in the introduction the central theme of both this and the final chapter are the algebraic relationships between the systems encountered in nonlinear filtering. We begin this section by discussing the origins of this algebraic estimation theory and outline the reasons justifying its existence.

The basic filtering problem we shall consider is the following. We suppose that a signal $\{x(t);t\geq 0\}$ is generated as the solution of the diffusion process in $\mathbb{R}^n$

$$(3.1.1)\qquad dx = f(x)dt + g(x)dw \qquad x(0) = x_0,$$

and that measurements of $x(t)$ are available through

$$(3.1.2)\qquad dy = h(x)dt + dv.$$

Here $f,g:\mathbb{R}^n\to\mathbb{R}^n$ and $h:\mathbb{R}^n\to\mathbb{R}$ are smooth functions and the noise processes $\{w(t)\}$, $\{v(t)\}$ are independent scalar Brownian motion.

Some comments are in order regarding the meaning of a solution to (3.1.1) and (3.1.2). Generally, these equations will be interpreted in the Ito sense. However, we can also interpret them through the pathwise constructions of Sussmann [4] and Doss [1]. Thus a solution of (3.1.1) is a stochastic process $\{x(t,\omega)\}$ defined on a probability space Ω, s.t. $\forall\ \omega\varepsilon\Omega$ the corresponding sample path satisfies the deterministic equation

$$(3.1.3) \qquad dx_{\omega}(t) = f(x_{\omega})dt + g(x_{\omega})dw_{\omega}(t)$$

Of course, $w_{\omega}(t)$ is only a continuous function so we also need a definition of a solution to (3.1.3) in this case. Such a definition is provided as follows. A curve $x:I\to\mathbb{R}^n$ defined on some interval $I\subset\mathbb{R}$, is a solution of (3.1.3) if $\exists$ a nhd U_{ω} of $w_{\omega}(.)$ in $C^0(I;\mathbb{R})$, the space of continuous functions from I into $\mathbb{R}$, and a continuous map $\Gamma:U\to C^0(I;\mathbb{R}^n)$ such that

$$\text{(i) } \Gamma(w_{\omega}) = x$$

and

$$\text{(ii) } \forall\ \bar{w}\varepsilon U_{\omega} \cap C^1(I;\mathbb{R}), \text{ then } \bar{x} \triangleq \Gamma(\bar{\omega}) \text{ satisfies the o.d.e.}$$

$$\frac{d\bar{x}}{dt} = f(\bar{x}) + g(\bar{x})\frac{d\bar{\omega}}{dt}$$

It turns out that the solution as defined here coincides with that of the Fisk-Stratonovich representation of (3.1.1) under suitable regularity conditions on f,g provided w is a scalar Brownian motion. This definition breaks down (without further conditions such as independence of noise, or commutativity of the corresponding input vector fields being imposed) for vector noise processes but note that this concept of a solution does carry through for any system of equations with continuous (sample) inputs. [We shall need two further observations regarding Fisk-Stratonovich integrals. Firstly, we recall that these can be obtained directly from Ito's definition of the stochastic integral resulting in the equivalent representation of (3.3.1) (đ denoting F.S. integration)

$$(3.1.3) \qquad \text{đ}x = \hat{f}(x)dt + g(x)\text{đ}\omega_t$$

with $\hat{f}(x) = f(x) - \frac{1}{2}\frac{dg}{dx}g$. Of more importance is the observation that (3.1.3) has the added advantage of satisfying the "usual rules of calculus".

In particular, if $\phi:\mathbb{R}^n\to\mathbb{R}$ is smooth enough then

$$\text{đ}\phi(x) = L_{\hat{f}}(\phi)dt + L_g(\phi)(x)\text{đ}\omega_t$$

whilst

$$d\phi(x) = \mathscr{L}(\phi)(x)dt + L_g(\phi)(x)d\omega_t$$

with $\mathscr{L}(\phi) = L_f(\phi) + \frac{1}{2}\Sigma g_i g_j \frac{\partial^2\phi}{\partial x_i \partial x_j}$.]

Now, given the system represented by (3.1.1) and (3.1.2) the objective of filtering is to obtain an estimate of $\psi(x(t))$ for some function $\psi:\mathbb{R}^n\to\mathbb{R}$ using only the information contained in the observations $\{y(s);\ 0 \leq s \leq t\}$. We denote this estimate by $e(\psi)(t)$. From a practical point of view (to ease implementation or computer storage, for instance) it is also desirable that this estimate be recursive, ie $\forall \Delta t > 0$ $e(\psi)$ satisfies

$$e(\psi)(t + \Delta t) = \Gamma(e(\psi)(t),\Delta t,\{y(s);t \leq s \leq t+\Delta t\})$$

so that the new estimate depends only on the new information and the old estimate. This recursiveness is often obtained by expressing $e(\psi)$ as the solution of a differential equation. We shall therefore say $e(\psi)$ is filterable if it satisfies

$$(3.1.4) \qquad \begin{cases} dz = a(z)dt + b(z)\text{đ}y \\ e(\psi)(t) = c(z(t)) \end{cases}$$

(Intuitively, one thinks of (3.1.4) as representing a 'black-box' with inputs the observations process and output $e(\psi)(t)$). Of course, the state space of (3.1.4) remains to be defined. If, in fact, it is a finite dimensional (smooth) manifold, so the dynamics are to be interpreted in the sense that $\forall$ smooth ϕ: $M\to\mathbb{R}$ we have

$$d\phi(z) = L_a(\phi)(z)dt + L_b(\phi)(z)\text{đ}y,$$

with a,b smooth vector fields and c a smooth function, then (3.1.4) is a (smooth) finite dimensional filter for ψ. In this case, ψ is then said to be a finite dimensionally computable (f.d.c) statistic. The central theme of algebraic estimation theory is the question of the existence of such f.d.c. statistics for the system defined by (3.1.1) and (3.1.2). It will be seen that this has fundamental links with the realisation theory of nonlinear systems.

Indeed, we certainly know that optimal estimates for any suitably regular statistic exist. For, by taking as our performance index the criterion of minimum variance, it is readily seen that if the underlying probability space is $(\Omega,\mathcal{A},P)$ and $\psi(x_t(.)) \in L_2(\Omega)$ then

$$e(\psi)(t) \overset{\Delta}{=} \hat{\psi}(t) = E(\psi(x_t)|\mathcal{Y}_t) \qquad (3.1.5)$$

where $\mathcal{Y}_t$ is the sub σ – algebra of $\mathcal{A}$ generated by $\{y(s);\ 0 \leq s \leq t\}$, is the required optimal estimate. Moreover, it turns out that $\hat{\psi}(t)$ is filterable and indeed several possible representations for filters exist (for an overview of results available in this area we refer to the excellent survey of Marcus and Davis [1], or for more detail to the texts of Kallianpur [1] or Lipster & Shiryaev [1]). Here, though, we consider an approach based on the unnormalised conditional density, $\rho(t,x)$, for the expectation in (3.1.5) so ρ is defined through the relation

$$p(t,x) = \frac{\rho(t,x)}{\int \rho(t,x)dx}$$

where $p(t,x)$ is the usual conditional density. The advantage to be gained by tackling the problem in this fashion is that ρ also satisfies the (conceptually) straightforward equation of Mortensen [1], Duncan [1], and Zakai [1] namely

$$d\rho = F(\rho)dt + G(\rho)dy \qquad (3.1.6)$$

where F and G are linear operators on $C^\infty(\mathbb{R}^n)$ defined by

$$F(\rho) = \tfrac{1}{2}\,\Sigma\frac{\partial^2}{\partial x_i \partial x_j}(g_i g_j \rho) - \Sigma\frac{\partial}{\partial x_i}(f_i \rho) - \tfrac{1}{2}h^2(x)\rho$$

$$G(\rho) = h(x)\rho$$

(for the systems we are considering, existence, uniqueness and regularity of the solution to (3.1.6) can be determined by the hypoellipticity of F, which in turn is guaranteed by the accessibility of the system $\dot{x} = f(x) + ug(x)$ c.f. Michel [1], Kunita [1], [2].) We therefore obtain our filter by augmenting (3.1.6) with the output

$$\hat{\psi}(t) = C_{\psi}(\rho)(t) \overset{\Delta}{=} \int_{\mathbb{R}^n} \psi(x)\rho(t,x)dx \left(\int \rho(t,x)dx\right)^{-1} \tag{3.1.7}$$

There remains a major obstacle to the use of this algorithm in any practical situation, namely the inherent infinite dimensionality of (3.1.6): ρ evolves in some function space, or, in more general descriptions, in a space of measures. It is natural, therefore, before attempting to implement this filter, to ask if there is not some simpler (preferably finite dimensional) description available. The Brockett Homomorphism Principle, Brockett [2], namely that, as a necessary condition for existence, there should exist a homomorphism from $\{F,G\}_{L.A.}$ onto a Lie algebra of vector fields on a finite dimensional manifold, is fundamental in this respect. This result has, as yet, only heuristic justification but it seems that only technical hypotheses obstruct a rigourous proof and some progress in this respect has recently been made by Hijab [1]. The basic argument is as follows. Suppose that a filter for the statistic $\hat{\psi}(t)$ exists in the desired form and is given by (3.1.4). Since the two representations are required to be equivalent for any data record or input, it is reasonable to assume that they are both realisations of the same stochastic input-output map, with "controls" having sample paths in $\mathcal{U} \subset C^0(\mathbb{R})$. From the previous discussions on the concept of the solution of a s.d.e., it is clear that this implies (recall we are assuming also that the stochastic integrals are in Fisk-Stratonovich form) that the underlying deterministic systems

$$\begin{cases} \dot{\rho} = F(\rho) + uG(\rho) & \rho(0) = \rho_0 \\ \hat{\psi} = C_{\psi}(\rho) \end{cases} \tag{3.1.8}$$

and

$$\begin{cases} \dot{z} = a(z) + ub(z) & z(0) = z_0 \\ \hat{\psi} = c(z) \end{cases} \tag{3.1.9}$$

also have the same input-output behaviour $\forall u \in \mathcal{U}$. The recent results of Hijab [1] further imply that (3.1.9) can be assumed to be minimal. Brockett's principle is then an immediate consequence of Theorem (1.2.5) - almost! The proof of Th^m (1.2.5) relies heavily on the differentiability of all the data. (3.1.8), however, is an infinite dimensional system and so more care must be taken (over domains etc). For these reasons we propose that (3.1.8) satisfies the following conditions.

(I) F and G are linear operators on a Banach space V and $\exists$ a subspace $D \subset V$ with $\rho_o \in D$ and any $X \in \{F,G\}_{L.A.}$ has a domain containing D. Moreover, D is invariant under both X and the (semi) flow generated by X.

(II) $C_\psi : V \to \mathbb{R}$ and for any analytic input, the output is also analytic (as functions of time).

Under these hypotheses, Brockett's Principle follows trivially by following the proof of Th^m(1.2.5.) (We remark that Brocketts original justification given in [2] was to assume the existence of a suitable generalisation of Sussmanns result (c.f. Th^m 1.2.4) that between any two finite dimensional realisations there was a map between the state spaces preserving trajectories, differentiation of which implied the associated Lie algebra homomorphism. Clearly, we have obtained our proof above independently of this assumption).

The existence of a domain D satisfying (i) above can also prove to be fundamental in obtaining a solution to (3.1.5). For, let us suppose that $\Lambda = \Lambda(\Sigma) = \{F,G\}_{L.A.}$ (henceforward to be referred to as the Estimation Algebra) is finite dimensional and defined on D. Then it is classical that Λ is isomorphic to the Lie algebra, $\mathfrak{g}$, of a unique Lie group G. Further, $\mathfrak{g}$ can be used to construct sets of coordinates on G through the exponential mapping. For suppose that $\{X_1,\ldots,X_n\}$ is a basis for $\mathfrak{g}$, and let $g_o \in G$. Then in a suitable neighbourhood U of g_o, every point can be reached by iteratively following suitable trajectories of $X_1 \ldots, X_n$. Thus, for some $\varepsilon>0$, $|t|<\varepsilon$ implies that for $g_1(t) \in U$ and $1 \leq i_j \leq n$, $1 \leq j \leq m$ we have

$$g_1(t) = \exp\eta_{i_1}(t)X_{i_1}(\ldots(\exp\eta_{i_m}(t)X_{i_m}(g_o))\ldots) \tag{3.1.10}$$

where $\frac{d}{dt}(\exp\eta_{ij}(t)X_{ij}(g)) = \dot\eta_{ij}(t)X_{ij}(\exp\eta_{ij}(t)X_{ij}(g))$

and η_{ij} are analytic functions satisfying $\eta_{ij}(0) = 0$ so $\exp(\eta_{kj}(0)X_{ij})(g) = g$.

Now assume that if $\pi_*:\mathfrak{g}\to\Lambda$ is the above mentioned isomorphism, then $\pi_*^{-1}(F) = Y_o$, $\pi_*^{-1}(G) = Y_1$ and $g_1(t)$ satisfies

$$\dot g_1(t) = Y_o + uY_1 \qquad g_1(0) = g_o \tag{3.1.11}$$

Differentiating, (3.1.10) we find

$$\dot g_1 = \dot\eta_{i_1}(t)X_{i_1}(g_1(t)) + \dot\eta_{i_2}(t)\exp(\eta_{i_1}X_{i_1})_* X_{i_2}(\exp\eta_{i_2}X_{i_2}(..)$$

$$+ \ldots.$$

Then, noting that for a diffeomorphism $\phi:G\to G$ and vector field X we have $\phi_*X(g) = \phi_*X(\phi^{-1}\phi(g))$, and using the Campbell-Baker-Hausdorff formula (2.1.1) we see

$$\dot g_1(t) = \eta_{i_1}X_{i_1}(g_1(t)) + \dot\eta_{i_2}\, e^{\mathrm{ad}_{\eta_{i_1}X_{i_1}}}(X_{i_2})(g_1(t))$$

$$+ \dot\eta_{i_z}\, e^{\mathrm{ad}_{\eta_{i_1}X_{i_1}}}\, e^{\mathrm{ad}_{\eta_{i_2}X_{i_2}}}(X_{i_2})(g_1) + \ldots.$$

and each expression involving $e^{\mathrm{ad}_Y}(X)$ is an element of $\mathfrak{g}$. Since $\{X_1,\ldots,X_n\}$ is a basis for $\mathfrak{g}$ and η_{i_j} are analytic, it follows that

$$\dot g_1(t) = \sum_{k=1}^{n} F_k(\dot\eta_1,\ldots,\dot\eta_m,\eta_1,\ldots,\eta_m)X_k \tag{3.1.12}$$

for some analytic functions F_k. Thus, a solution to (3.1.11) can be found by writing Y_o and Y_1 in terms of the basis vector fields, "equating coefficients" and solving the resulting o.d.e's for $\eta_i(t)$. {The technique described above is essentially due to Wei and Norman [1]}. At a formal level this analysis also works for the Zakai equation (3.1.6). However, in order to obtain a solution to equation (3.1.5) knowing the solution to (3.1.11) we really need to know that the representation π_* "integrates" to a representation of G on V. That is, we need to find a differentiable isomorphism $\pi : G \to GL(V)$ with tangent map π_* and such that the following diagram commutes:

$$\begin{tikzcd}
\mathfrak{g} \arrow[r, "\pi_*"] \arrow[d, "\exp"'] & \Lambda \arrow[d, "\exp"] \\
G \arrow[r, "\pi"'] & M \subseteq GL(V)
\end{tikzcd}$$

(π_* is then said to be the differential of π). This is a well-known problem in Lie algebra representation theory whose solution usually requires a further analyticity condition on D, coupled with an existence result for trajectories of the operators in Λ corresponding to the basis $\{X_1,..,X_n\}$ in $\mathfrak{g}$. Typical in this respect is the following result which reflects the essential features, but is included solely as an example of the genre (we also suggest Kirillov [1], Jorgenson [1], Moore [1], and Flato et al. [1] as further sources).

<u>THEOREM 3.1.1</u> (Simon [1])

Let $\mathfrak{g}$ be a finite dimensional Lie algebra with generators $\{x_1,...,x_n\}$ and suppose π_* is a representation of $\mathfrak{g}$ on a reflexive Banach space V with $D \subset V$ a dense domain for $\pi_*(\mathfrak{g})$ satisfying (I). Assume further that

a) if $X_i = \pi_*(x_i)$, then X_i^* (the dual of X_i) has a dense domain D_i^* of analytic vectors, (i.e. for all $v \in D_i^*$ the series $\sum\limits_{k \geq 0} \frac{t^k}{k!} X_i^{*k}(v)$ is absolutely convergent) such that $D_i^* \subset D_j^*$, $X_j^*(D_i^*) \subset D_i^*$ and $\overline{X_j^*|_{D_i^*}} = X_j^*$, $1 \leq i \leq n$, $1 \leq j \leq n$, $\overline{\cdot}$ denoting closure, and

b) the operators X_j^* generate strongly continuous one parameter groups on V^*.

Then π_* is the differential of a unique representation of the corresponding Lie group on V.

□

In most situations regarding exponentiation of the estimation algebra, the problem will deviate from the set up described in Th^{m} (3.1.1), for, as pointed out in Brockett [3], it will often be the case that the operators in Λ will not have trajectories defined for all time, and thus fail to satisfy condition b). However, the essential technical point is the analyticity condition on the domains, which allows the local construction of solutions. We illustrate the technique outlined above by considering the simple linear system

$$(3.1.13) \qquad \begin{aligned} dx &= dw \\ dy &= x\,dt + dv \end{aligned}$$

for which the estimation algebra is generated by

$$F = \tfrac{1}{2}\frac{\partial^2}{\partial x^2} - \tfrac{1}{2}x^2 \quad \text{and} \quad G = x$$

Hence, $\Lambda = \mathrm{Sp}\{F, G, \partial/\partial x, 1\}$ and has commutation table

	F	G	$\frac{\partial}{\partial x}$	1
F	0	$\frac{\partial}{\partial x}$	G	0
G	$-\frac{\partial}{\partial x}$	0	-1	0
$\partial/\partial x$	$-G$	1	0	0
1	0	0	0	0

From which we see that Λ is actually solvable. In Wei-Norman [1], it is shown that this is actually a sufficient condition, again assuming suitable integrability, for the solution to the corresponding version of (3.1.5) to be given by

$$(3.1.14) \quad \rho(t,x) = \exp(\eta_1(t)X_1(\exp\eta_2(t)X_2(\exp\eta_3(t)X_3(\exp\eta_4(t)X_4\ \rho(0))..)$$

$\forall t \in \mathbb{R}^{+}$, where $\{X_1, .., X_4\}$ is an ordered basis for Λ. This basis is determined using Lie's Theorem, and for this example is given by

$$X_1 = F,\ X_2 = G - \frac{\partial}{\partial x},\ X_3 = G + \frac{\partial}{\partial x},\ X_4 = 1.$$

We now proceed formally by differentiating (3.1.14). This gives

$$\frac{\partial\rho}{\partial t} = (\dot{\eta}_1 X_1 + \{\dot{\eta}_4 - 2\dot{\eta}_3\eta_2\}\, X_4 + \dot{\eta}_2 e^{-\eta_1} X_2 + \dot{\eta}_3 e^{\eta_1} X_3(\rho)$$

$$= (F + uG)(\rho)$$

Hence

$$(3.1.15)\quad \begin{cases} \dot{\eta}_1 = 1 \\ \dot{\eta}_2 e^{-\eta_1} + \dot{\eta}_3^{\eta_1} = u(t) \\ \dot{\eta}_4 - 2\dot{\eta}_3\eta_2 = 0 \\ \dot{\eta}_2 e^{-\eta_1} - \dot{\eta}_5^{-\eta_1} = 0 \end{cases} \qquad \eta_i(0) = 0 \quad 1 \leq i \leq 4.$$

We now define $V = \mathrm{Sp}\{x^{\alpha} e^{\beta x}\psi ; \psi \varepsilon L^1(\mathbb{R}), \alpha, \beta \varepsilon \mathbb{R}\}$. Then it is shown in Ocone [1], that $V \cap C^{\infty}(\mathbb{R})$ is an analytic domain for this problem. Moreover, on V

$$(3.1.16)\quad \begin{cases} \exp\{\eta_2 X_2\}(\phi)(x) = e^{(\eta_2 x - \eta_2^2/2)}\phi(x-\eta_2) \\ \exp\{\eta_3 X_3\}(\phi)(x) = e^{(\eta_3 x + \eta_2^2/2)}\phi(x+\eta_3) \\ \exp\{\eta_4 X_4\}(\phi)(x) = e^{\eta 4}\phi(x) \\ \exp\{tX_1\}(\phi)(x) = \int_{\mathbb{R}} G(x,t,y)\phi(y)\,dy \end{cases}$$

are the trajectories generated by the linear operators $\phi \to \eta_i X_i(\phi)$ for $\phi \varepsilon \mathcal{D}$, $1 \leq i \leq 4$, and $G(x,t,y)$ is the Greens' function

$$G(x,t,y) = \frac{1}{(2\sinh t\pi)^{\frac{1}{2}}} \exp\{-\tfrac{1}{2}\coth t\,(x^2+y^2) + \frac{xy}{\sinh t}\}$$

[Note that $\exp\{tX_1\}: L^1(\mathbb{R}) \to C^{\infty}(\mathbb{R})\ \forall t>0$, so that (3.1.14) as determined by (3.1.15) and (3.1.16) is a smooth solution to the unnormalised conditional density equation $\forall L^1$ initial condition and suitable inputs]. We do not solve these equations here, contenting ourselves with remarking that the Kalman Filter for (3.1.13)

$$\begin{cases} d\hat{x} = p\,\hat{x}\,dt + p\,dy \\ \dot{p} = 1-p^2 \end{cases}$$

is readily obtained. We also observe that the pair (3.1.15), (3.1.14) together with (3.1.7) serves to define a further finite dimensional filter, since we can write $\rho(t,x) = \Gamma(\eta_1,..,\eta_4;\rho_o(x))$ so

$$C_\psi(\rho(t,x)) = C_\psi(\Gamma(\eta_1,\eta_2,\eta_3,\eta_4,\rho_o(x))) = \bar{h}_\psi(\eta)$$

and thus

$$\text{(3.1.17)} \quad \left|\begin{array}{l} \dot{\eta} = \begin{bmatrix} \dot{\eta}_1 \\ \cdot \\ \cdot \\ \cdot \\ \dot{\eta}_4 \end{bmatrix} = \bar{f}(\eta) + u\bar{g}(\eta) \qquad \eta(0) = 0 \\ \hat{\psi}(t) = \bar{h}_\psi(\eta) \end{array}\right.$$

with $\bar{f}(\eta) = \frac{\partial}{\partial \eta_1}$, $\bar{g}(\eta) = (\frac{1}{2}e^{\eta_1}\frac{\partial}{\partial \eta_2} + \frac{1}{2}e^{-\eta_1}\frac{\partial}{\partial \eta_3} + e^{-\eta_1}\eta_2\frac{\partial}{\partial \eta_4})$ is the required filter. It is also readily seen that the linear extension of the map $F \to \bar{f}$, $G \to \bar{g}$, $1 \to \frac{\partial}{\partial \eta_4}$, $\frac{\partial}{\partial x} \to (\frac{1}{2}e^{\eta_1}\frac{\partial}{\partial \eta_2} - \frac{1}{2}e^{-\eta_1}\frac{\partial}{\partial \eta_3} - \eta_2 e^{-\eta_1}\frac{\partial}{\partial \eta_4})$ is a Lie algebra isomorphism between the estimation algebra and $\{\bar{f},\bar{g}\}_{L.A.}$ A similar construction is also possible for the Kalman Filter representation. A point of further interest is that although (3.1.17) is clearly accessible, for the case that $\hat{\psi}(t) = \hat{x}(t) = E(x|\mathcal{Y}_t)$, it cannot be minimal (the Kalman Filter is defined on $\mathbb{R}^2$ whilst (3.1.17) is defined in $\mathbb{R}^4$) and hence cannot be observable.

The preceding discussions show the importance of the finite dimensionality of the Estimation Algebra - indeed this condition not only provides an immediate answer to the central question of algebraic estimation, namely when does the Estimation Algebra satisfy Brocketts Homomorphism Principle, but also gives insight into the possible subsequent construction of a finite dimensional filter. For, as a direct consequence of Ado's theorem (stating that any finite dimensional Lie algebra is isomorphic to a Lie algebra of matrices), it follows that if $\Lambda(\Sigma)$ is finite dimensional then the dynamic equation (3.1.6) can be represented in a bilinear form

(3.1.18) $$d\xi = A\xi\, dt + B\xi dy \qquad \xi \in \mathbb{R}^n,\ A,B \in g\ell(n,\mathbb{R})$$

such that $\{A,B\}_{L.A.} \cong \Lambda(\Sigma)$. Of course, this does not necessarily mean the input-output map for the corresponding statistic can be realised on $\mathbb{R}^n$. For this to be the case, one also needs to find a suitable output function and initial condition for (3.1.18). We shall return to this point, briefly, in § 3.2. For the remainder of this section, however, we concentrate on trying to establish when the Estimation Algebra is finite dimensional. As will be seen, this has intimate connections with the system algebra and observation space of the underlying deterministic system associated with (3.1.1) and (3.1.2).

Our initial observation, as a starting point for this investigation, is to note that after some algebraic manipulation the generator F as defined in (3.1.6) can be written in the form

(3.1.19) $$F\rho = \tfrac{1}{2}L^*_g(L^*_g(\rho)) - L^*_{\hat{f}}(\rho) - \tfrac{1}{2}h(x)^2 = \tfrac{1}{2}L^{*2}_g(\rho) - L^*_{\hat{f}}(\rho) - \tfrac{1}{2}h^2(x)\,\rho$$

where, for a vector field X on $\mathbb{R}^n$, L^*_X is the formal adjoint of L_X. Thus

$$L^*_X(\phi)(x) = -\Sigma \frac{\partial}{\partial x_i} X_i(x)\phi(x) \qquad \forall \phi \in C^\infty(\mathbb{R}^n)$$

(note, L^*_X is the natural extension of the adjoint of L_X defined on $C^\infty_0(\mathbb{R}^n)$ with respect to the standard L_2 inner product). $\hat{f}$ is the perturbed, or Ito corrected, version of f defined in (3.1.3). The following lemmas will prove useful in the subsequent Lie algebraic calculations.

LEMMA 3.1.2

Let A be associative algebra and let $[.,.]$ be the standard commutator bracket on A, Then, $\forall$ X,Y,Z$\in$A

a) $[X,YZ] = [X,Y]Z + Y[X,Z]$

b) $[X,Y^2] = 2Y[X,Y] + [[X,Y],Y]$

$\qquad\qquad = 2[X,Y]Y - [[X,Y],Y]$

Proof

Both identities are the result of trivial algebra. For a) we need only expand the R.H.S.

$$[X,Y]Z + Y[X,Z] = \{XYZ - YXZ\} + YXZ - YZX$$
$$= XYZ - YZX = [X,YZ].$$

b) follows directly from a) and the definition of the commutator.

□

LEMMA 3.1.3

If ϕ,ψ are C^∞ functions on $\mathbb{R}^n$ and $X,Y\varepsilon\Gamma^\infty(T\mathbb{R}^n)$ then

a) $L_X^*(\phi\psi) = \phi L_X^*(\psi) - L_X(\phi)$

b) $L_Y^*L_X^*(\phi\psi) = \phi L_Y^*L_X^*(\psi) - L_Y(\phi)L_X^*(\psi) - L_Y^*(\psi)L_X(\phi) + \psi L_Y L_X(\phi).$

Proof

Again b) follows directly from a) whilst a) itself is an immediate consequence of the definition of the adjoint

$$L_X^*(\phi\psi) = -\Sigma \frac{\partial}{\partial x_i} X_i\phi\psi = -\Sigma \frac{\partial}{\partial x_i} \phi(X_i\psi)$$
$$= - \{\Sigma X_i\psi \frac{\partial\phi}{\partial x_i} + \phi\frac{\partial}{\partial x_i} (X_i\psi)\}$$
$$= - L_X(\phi)\psi + \phi L_X^*(\psi)$$

□

Next notice that $\{F,G\}_{L.A.}$ considered as a Lie algebra of vector fields on some analytic domain D, is isomorphic to $\{F,G\}_{L.A.}$ considered as differential operators on $\mathbb{R}^n$. The proof of this fact is identical in all qualitative respects to the similar result that for a bilinear system the associated Lie algebra is isomorphic to the Lie algebra generated by the matrices defining the dynamics. More precisely, suppose that Λ is defined on a domain D and $\psi:D\to\mathbb{R}$ is differentiable. Then $\forall X\varepsilon\Lambda$ the Lie derivative is defined by

$$L_X(\psi)(\rho) = \frac{d}{dt} \psi(\gamma_t^X(\rho))\Big|_{t=0}$$

with $\gamma_t^X(\rho)$ a local trajectory for X through ρ. Assuming sufficient structure on D, this can be written

$$L_X(\psi)(\rho) = D\psi_\rho \; X(\rho)$$

Then

$$\begin{aligned} L_X L_Y(\psi)(\rho) &= D(L_Y\psi)_\rho X(\rho) \\ &= D^2\psi_\rho(Y(\rho),X(\rho)) + D\psi_\rho DY_\rho \quad X(\rho) \end{aligned}$$

But Y is a linear operator, so $DY_\rho X(\rho) = YX(\rho)$, and the 2nd derivative drops out on taking the Lie Bracket since $D^2\psi_\rho$ is a symmetric bilinear mapping. Thus,

$$L_{[X,Y]}(\psi)(\rho) = -D\psi_\rho.[Y,X](\rho)$$

and the isomorphism claim follows trivially

The significance of Lemma (3.1.3) now becomes clear, for as immediate corollaries we have, $\forall\phi,\psi\varepsilon C^\infty(\mathbb{R}^n)$

$$[L^*_{X_1},\phi](\psi) = -L_{X_1}(\phi)(\psi)$$

and using (3.1.2)(b)

$$\begin{aligned} [L^{*2}_{X_2},\phi](\psi) &= [L^*_{X_2},[L^*_{X_2},\phi]]\;(\psi) + 2[L^*_{X_2},\phi]L^*_{X_2}\,(\psi) \\ &= L^2_{X_2}(\phi)\psi - 2L_{X_2}(\phi)L^*_{X_2}(\psi) \end{aligned}$$

In particular, from the form of F given in (3.1.19) and assuming h is a C^∞ function we see

$$\begin{aligned} [F,G](\psi) &= -\,L_g(h)L^*_g(\psi) + (L_{\hat{f}}(h) + \tfrac{1}{2}L^2_g(h))(\psi) \\ &= -\,L_g(h)L^*_g(\psi) + \mathscr{L}(h)(\psi) \end{aligned}$$

where $\mathscr{L} \triangleq L_{\hat{f}} + \frac{1}{2}L^2_g$ is the Fokker Planck operator associated with (3.1.1) and (3.1.2). Next we see that since G acts as a multiplication operator it will commute with any other such mapping. In other words, $\forall\phi\varepsilon C^\infty(\mathbb{R}^n)$, $[G,\phi] = 0$ so we first see that

$$\begin{aligned}[G,[F,G] &= -[G, L_g(h)L_g^* - \mathscr{L}(h)] \\ &= -[G, L_g(h)L_g^*] \\ &= L_g(h)^2\end{aligned}$$

and hence $[G,[G,[F,G]]] = 0$. Again using Lemma (3.1.2) we find

$$\begin{aligned}\mathrm{ad}_F^2(G) &= [F, \mathscr{L}(h) - L_g(h)L_g^*] \\ &= [F,\mathscr{L}(h)] - [F,L_g(h)]L_g^* - L_g(h)[F,L_g^*]\end{aligned}$$

Clearly, the brackets in the first two terms of the r.h.s. of this equation are determined in exactly the same way as $[F,G](=[F,h])$ was calculated. Also noting that $[L_g^{*2}, L_g^*] = 0$ it then takes straightforward manipulation to show

$$\mathrm{ad}_F^2(G) = L_g^2(h)L_g^{*2} - \{L_g\mathscr{L}(h) + \mathscr{L}L_g(h)\}L_g^* + L_g(h)[L_f^*, L_g^*] + L_g(h)^2 h + \mathscr{L}^2(h)$$

Without more specific knowledge of the system under consideration the operators $\mathrm{ad}_F^k G$ rapidly become complicated objects. However, some structural properties can be observed. As we have already pointed out $\Lambda(\Sigma)$ is a Lie algebra of C^∞ differential operators on $\mathbb{R}^n$. Let us denote by $\mathcal{D}^\infty$ ($\mathcal{D}_k^\infty$) the vector space of all such operators (resp. those of degree k+1). Thus,

$$\mathcal{D}_k^\infty = \{X \in \mathcal{D}^\infty;\ X = \sum_{0 \leq |\alpha| \leq k+1} X_\alpha(x) \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha} \ldots \partial x_n^{\alpha_n}}, X_\alpha \in C^\infty(\mathbb{R}^n)\}$$

It is a straightforward exercise, using Leibnitz' formula, to see that $\{\mathcal{D}_k; k \geq -1\}$ defines a filtration on $\mathcal{D}^\infty$, with $\mathcal{D}_{-1}^\infty$ $(= C^\infty(\mathbb{R}^n))$ and $\mathcal{D}_0^\infty$ Lie subalgebras of $\mathcal{D}^\infty$, which naturally induces a similar filtration $\{\Lambda_k\}$ on $\Lambda(\Sigma)$ with

$$\Lambda_k \triangleq \Lambda(\Sigma) \cap \mathcal{D}_k^\infty \qquad k \geq -1$$

Now, since we can write

$$L_g^{*2} = \sum_{j,i=1}^{n} \{g_i g_j \frac{\partial^2}{\partial x_i \partial x_j} + g_i \frac{\partial g_j}{\partial x_i} \frac{\partial}{\partial x_j}\}$$

and $\{\frac{\partial^2}{\partial x_i \partial x_j}, \frac{\partial}{\partial x_k}\ ;\ 1 \leq i,j,k \leq n\}$ are all linearly independent over $C^\infty(\mathbb{R}^n)$, it follows that $F \in \Lambda_1$ provided $g \not\equiv 0$ and, by definition $G \in \Lambda_{-1}$. In general, then, we can draw two conclusions from the algebraic structure of Λ. First, G is a locally nilpotent operator, ie $\forall X \in \Lambda \, \exists \, n \geq 0$ s.t. $\mathrm{ad}_G^n(X) \equiv 0$. For, if $X \in \Lambda$, then $X \in \Lambda_{n-1}$ for some minimal n. Since $[\Lambda_{-1}, \Lambda_k] \subset \Lambda_{k-1} \; \forall k$, and $G \in \Lambda_{-1}$ the claim follows immediately. In a similar vein, it is easy to see that if $X \in \Lambda_k$ then $\mathrm{ad}_F(X) \in \Lambda_{k+1}$. Both of these properties are evidenced in the above calculations of $\mathrm{ad}_F G$ and $\mathrm{ad}_F^2 G$.

As a corollary of particular interest, notice that each generator $\mathrm{ad}_F^k G \in \Lambda_{k+1}$. Moreover, the principal part of $\mathrm{ad}_F^k G$ is readily seen to be determined by $L_g^k(h) L_g^{*k}$ in the sense that

$$\mathrm{ad}_F^k G = (-1)^k L_g{}^k(h) L_g^{*k} + Y_k \qquad \text{for some } Y_k \in \mathcal{D}_k^\infty \tag{3.1.20}$$

The validity of this comment is easily established through a simple induction using the previous lemmas and the observation that ad_F is only 'degree increasing' due to presence of the L_g^{*2} term. Thus,

$$\mathrm{ad}_F^{k+1} G = \tfrac{1}{2}[L_g^{*2}, \mathrm{ad}_F^k G] - [L_f^* + \tfrac{1}{2}h^2, \mathrm{ad}_F^k G]$$

The second bracket can have degree at most k since $(L_f^* - \tfrac{1}{2}h^2) \in \mathcal{D}_0^\infty$. Similarly, $[L_g^{*2}, Y_k] \in D_{k+1}^\infty$, so that the $(k+1)^{\text{th}}$ order element of $\mathrm{ad}_F^{k+1} G$ is given as

$$\begin{aligned} \frac{(-1)^k}{2}[L_g^{*2}, L_g{}^k(h) L_g^{*k}] &= \frac{(-1)^k}{2}[L_g^{*2}, L_g^k(h)] L_g^{*k} \\ &= (-1)^{k+1} L_g{}^{k+1}(h) L_g^{*k+1} + \frac{(-1)^k}{2} L_g{}^{k+2}(h) L_g^{*k} \end{aligned}$$

and, hence, proves the claim.

This result gives an immediate necessary condition for the estimation algebra to be finite dimensional. We define a sequence of subspaces $V_n \subset \Lambda(\Sigma)$ by setting

$$V_n = \mathrm{Sp}\{\mathrm{ad}_F^k G \; ; \; k \in \{0, \ldots, n\}\}.$$

If we assume Λ is finite dimensional, then by the ascending chain condition it follows that this sequence must have a maximal element and so there is an integer k for which $\dim V_n = \dim V_k \; \forall n \geq k$. In particular, we then find that

$$(3.1.21) \qquad \mathrm{ad}_F^{k+1} G = \sum_{j=0}^{k} \alpha_j \mathrm{ad}_F^j(G) \qquad \alpha_j \in \mathbb{R}$$

However, $\mathrm{ad}_F^{k+1} G \in \Lambda_{k+2}$, whilst the r.h.s. of (3.1.21) is in Λ_{k+1}. From (3.1.20) it therefore follows that $L_g^{k+1}(h) L_g^{*k} \equiv 0$. But L_g^{*k} is a non-zero operator, for non zero g, so assuming further that g and h are analytic, we must have $L_g^{k+1}(h) \equiv 0$. Somewhat surprisingly, this condition can be strengthened considerably.

THEOREM 3.1.4

Consider the system (3.1.1) and (3.1.2) and assume that $g \in \Gamma^{\omega}(T\mathbb{R}^n)$ $h \in C^{\omega}(\mathbb{R}^n)$ and $\Lambda(\Sigma)$ is finite dimensional. Then $L_g^2(h)$ is constant.

Proof

Let $k = \min\{\ell; L_g^{\ell}(h) = 0\}$ and set $Z_1 = \mathrm{ad}_F^{k-1} G$. Now consider $X_1 = [Z_1, \mathrm{ad}_F G]$. From the filtration properties of $\Lambda(\Sigma)$ if follows that both Z_1 and $X_1 \in \Lambda_k$ and using the same arguments as those used to prove (3.1.20) we see that the principal part of X_1 is given by

$$\begin{aligned} p(X_1) &= p([L_g^{k-1}(h) L_g^{*k-1}, L_g(h) L_g^{*}]) \\ &= p\{(k-1) L_g^{k-1}(h) L_g^2(h) - L_g(h) L_g^k(h)\} L_g^{*k-1}) \\ &= L_g^{k-1}(h) L_g^2(h) L_g^{*k-1} . \end{aligned}$$

(In this context, the principal part of a differential operator we mean its highest order terms with one coefficient normalised). Inductively, we define two sequences in Λ by the recursions

$$\begin{aligned} Z_{n+1} &= \mathrm{ad}_F^{k-3}(X_n) \qquad\qquad n \geq 1 \\ X_n &= [Z_n, \mathrm{ad}_F G] \end{aligned}$$

We claim that $p(Z_n) = (L_g^{k-1}(h))^n L_g^{*nk-3n+2}$. For n=1 this is true by definition, so we assume it to be also true for n=1,...,N. Then, it is easy to see

$$\begin{aligned} p(X_N) &= p([(L_g(h))^N L_g^{*Nk-3N+2}, L_g(h) L_g^{*}]) \\ &= (L_g^{k-1}(h))^N L_g^2(h) L_g^{*Nk-3N+2} \end{aligned}$$

and

$$p(Z_{n+1}) = p(\mathrm{ad}^{k-3}_{L_g^{*2}}((L_g^{k-1}(h)^N L_g^2(h) L_g^{*Nk-3N+2}))$$

$$= L_g^{k-1}(h))^{N+1} L_g^{*(N+1)(k-3)+2}$$

as required. Thus Z_n is of increasing order, and hence a linearly independent sequence, unless

$$(N+1)(k-3) = N(k-3)$$

ie $k = 3$ and $L_g^3(h) = 0$. But then,

$$Z_n = L_g^2(h)^N L_g^{*2} + \text{l.o.ts},$$

which still gives an infinite linearly independent sequence unless $L_g^2(h)$ is constant. This proves the theorem.

□

We remark that this result has not been stated in its full generality. Since we have only used the properties of ad_F acting on functions, we can easily adapt the above proof to show that the following result is also true.

THEOREM (3.1.4a)

Let V be a finite dimensional subspace of $\Lambda(\Sigma)$ which is ad_F-invariant. As before assume that Σ is a linear analytic system. Then $\phi \epsilon V \cap \Lambda_{-1} \Rightarrow L_g^2(\phi)$ is constant. (We only need to check that any such ϕ is also analytic, but this is obvious since it can only be generated by a sequence of multiplications by analytic functions or Lie derivations by analytic operators, from h).

□

In the next section, this theorem is used to analyse scalar polynomial systems and a limited class of multi-input systems, so for the present we merely note the condition is trivially satisfied by linear systems, in which case g is a constant vector field and h is a linear function. However, it is easy to see that the criterion is not sufficient. Indeed, there is a well-known example, due to Hazewinkel-Marcus [1], of a simple bilinear system on $\mathbb{R}^2$ which satisfies $L_g^2(h)$ = constant but whose estimation algebra is infinite dimensional and contains no ideals isomorphic to a Lie algebra of vector fields on a finite dimensional manifold. This example will be presented in, and forms the basis of, the final chapter.

Having established that there is a connection between the input-vector field, g, the output function h and finite dimensionality of the Estimation Algebra, we next turn our attention to the role of the drift field.

THEOREM (3.1.5)

Assume (3.1.1) and (3.1.2) define a linear analytic system on $\mathbb{R}^n$. Then

a) if $\hat{f} \equiv 0$, $\Lambda(\Sigma)$ is finite dimensional $\Leftrightarrow$ $L_g(h)$ is constant

b) if $L^*_{\hat{f}} \in \Lambda(\Sigma)$ and $\Lambda(\Sigma)$ is finite dimensional, then $L_g(h)$ is constant

c) if $h^2 \in \Lambda(\Sigma)$ and $\Lambda(\Sigma)$ is finite dimensional, then $L_g(h)$ is constant.

Proof

a) if $\hat{f} \equiv 0$ then $F = \frac{1}{2}\{L_g^{*2} - h^2\}$. We first assume $\dim\Lambda(\Sigma) < \infty$. Then by Th^m (3.1.4) we must have $L_g^2(h) = c$ so that

$$[F,G] = -L_g(h)L_g^* + \tfrac{1}{2}c$$

and

$$ad_F^2 G = cL_g^{*2} + L_g(h)^2 h.$$

Assume that $c \neq 0$. Then

$$ad_F^2 G = 2cF = ch^2 + hL_g(h)^2 \triangleq \phi_o$$

From Th^m (3.1.4a) we then see that $L_g^3(\phi_o) = 0$. However trivial calculation yields

$$L_g^3(\phi_o) = \alpha c^2$$

and $\alpha \in \mathbb{R}$ is non zero. Thus $c = 0$. Consequently,

$$ad^2{}_F G = L_g(h)^2 h$$

Inductively, it is not difficult to see that

$$ad^{2n}{}_F G = L_g(h)^{2n} h$$

so as in Th^m (3.1.4), dim $\Lambda < \infty$ forces $L_g(h)$ to be constant. The converse implication is just as straightforward. If $L_g(h) \equiv 0$, the result is trivial since $[F,G] = 0$ so $\Lambda = \mathrm{Sp}\{F,G\}$. Assume therefore that $L_g(h) = c \neq 0$. Then

$$[F,G] = -cL_g^* \quad \Rightarrow \quad L_g^* \in \Lambda$$

and

$$[F,[F,G]] = L_g(h)^2 h = ch$$

Thus,

$$\begin{aligned}\Lambda &\triangleq \mathbb{R}F + \{ad_F^k G;\ k \geq 0\}_{L.A.} \\ &= \mathbb{R}F + \{L_g^*, h\}_{L.A.} \\ &= \mathrm{Sp}\{F, L_g^*, h, 1\}.\end{aligned}$$

In particular dim $\Lambda < \infty$

b) This follows trivially using the same proof as that of a) after noting that since $L_f^* \in \Lambda$, we also have $L_g^{*2} - h^2 \in \Lambda$.

c) Since $h^2 \in \Lambda$, this result follows directly from Th^m (3.1.4a), from which

$$L_g^3(h^2) = 6L_g(h)L_g^2(h) \equiv 0$$

However, $L_g(h), L_g^2(h) \in C^\omega(\mathbb{R}^n)$ so that either $L_g(h) \equiv 0$ or $L_g^2(h) \equiv 0$. Clearly, if $L_g(h) = 0$, then $L_g^2(h) = 0$ so in either case we must have $L_g^2(h) = 0$

If we now set $\phi_o = h^2$, and define $\phi_{n+1} = [ad_F(\phi_n), \phi_o]$ it is easy to see that

$$\phi_n = L_g(h)^{2n} h \quad \in \Lambda \ \forall n \geq 0$$

From which, using the same arguments as before, we readily infer that $L_g(h)$ must be constant if dim $\Lambda < \infty$. □

Possibly the simplest example of a system satisfying part a) of this theorem non-trivially is that considered in (3.1.13). In fact, it is not difficult to see that the commutation relations for $\{F, L_g^*, h, 1\}_{L.A.}$ are identical to those given for $\{\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2, \frac{\partial}{\partial x}, x, 1\}_{L.A.}$ so that these Lie algebras are isomorphic. This is not surprising, for if we consider the deterministic system,

$$
(3.1.21) \qquad \begin{cases} \dot{x} = ug(x) \qquad x \in \mathbb{R}^n \\ \\ y = h(x) \end{cases}
$$

satisfying $L_g(h) = 1$ then clearly

$$\dot{y} = u(t) L_g(h)(x) = u(t)$$

So that (3.1.21) is merely a (non-minimal) realisation of the underlying system

$$
\begin{cases} \dot{\xi} = u \\ y = \xi \end{cases}
$$

of (3.1.13).

As a final remark on this theorem, note that part c) takes on added significance when it is realised that if $\Lambda_1 \triangleq \{\frac{1}{2} L_g^{*2} - L_{\hat{f}}^{*}, h^2 . h\}_{L.A.}$ then a sufficient condition for the estimation algebra to be finite dimensional is that Λ_1 be finite dimensional. This follows trivially from the observation that $F \in \Lambda_1$ so that $\Lambda \subset \Lambda_1$ (in fact, if $h^2 \in \Lambda$, then $\Lambda_1 \subset \Lambda$). A necessary condition for Λ_1 to be finite dimensional is then that $L_g^2(h^2)$ be constant: the proof of Th^m (3.1.4) readily adapting to this new situation, since the presence of L_g^{*2} still causes the same degree increase.

The most interesting point to note about these two theorems is that the conditions derived are essentially restrictions on the observation space of the system $\hat{\Sigma} = \{\hat{f}, g, h\}$. To conclude this section we present a general containment result which goes some way towards explaining this phenomenon. The notation Σ^* will refer to the 'system' $\{L_f^*, I_g^*, h\}$.

THEOREM 3.1.6

For any linear analytic system, $\Sigma_1 = \{f, g, h\}$ we have

$$\Lambda(\Sigma) \subset \mathbb{R} F + H_A(\hat{\Sigma}) \otimes \mathscr{U}(\mathscr{S}(\hat{\Sigma}^*)) \triangleq \Omega(\hat{\Sigma})$$

where $\mathscr{U}(\mathfrak{g})$ is the universal enveloping algebra of the Lie algebra $\mathfrak{g}$ so

$$\mathscr{U} = \bigoplus_{j \geq 0} \mathfrak{g}^{\otimes j}$$

Proof

Clearly, F and G $\epsilon\Omega(\Sigma)$ so we only need to prove that $\Omega^o(\hat{\Sigma}) \overset{\Delta}{=} H_A(\hat{\Sigma}) \theta \mathscr{U}(\mathscr{S}(\hat{\Sigma}^{**}))$ is an ad_F-invariant Lie subalgebra of $\Omega(\hat{\Sigma})$. Now by definition

$$X\epsilon\Omega^o \Rightarrow X = \sum_{0\leq|\alpha|\leq k} \phi_\alpha X_1^{\alpha_1}\cdots X_n^{\alpha_n}$$

for some $X_1,\ldots,X_n \epsilon \mathscr{S}(\hat{\Sigma}^{**})$, $\phi_\alpha \epsilon H_A(\hat{\Sigma})$ and $n\geq 0$. Then, from Lemma (3.1.2) we find

$$\mathrm{ad}_F(X) = \Sigma \mathrm{ad}_F(\phi_\alpha)X_1^{\alpha_1}..X_n^{\alpha_n} + \phi_\alpha \, \mathrm{ad}_F(X_1^{\alpha_1}..X_n^{\alpha_n}) \tag{3.1.22}$$

Since $\mathrm{ad}_F(\phi_\alpha) = -L_g(\phi_\alpha)L_g^* + \mathscr{L}(\phi_\alpha)$ the first term in this expansion is certainly in Ω^o. Similarly, if $X_i \epsilon \mathscr{S}(\hat{\Sigma}^{**})$

$$\begin{aligned} \mathrm{ad}_F(X_i) &= [L_g^*, X_i]L_g^* + \tfrac{1}{2}[L_g^*,[L_g^*,X_i]] - [L_f^*, X_i] - \tfrac{1}{2}[h\ ,X_i] \\ &\epsilon\ \mathscr{S}(\hat{\Sigma}^{**})\ \theta\ \mathscr{S}(\Sigma^*) + \mathscr{S}(\Sigma^*) + H_A(\hat{\Sigma}) \\ &\subset\ \Omega^o(\Sigma) \end{aligned}$$

Inductive use of Lemma (3.1.2) shows that the second sum in (3.1.22) is also in $\Omega^o(\Sigma)$ and hence $\mathrm{ad}_F(\Omega^o) \subset \Omega^o$.

Now suppose that $Y\epsilon\Omega^o$ and $Y = \sum_{|\beta|\leq j} \psi_\beta Y_1^{\beta_1}..Y_n^{\beta_n}$. Then

$$[X,Y] = \sum_{\substack{|\alpha| \\ |\beta|}} [\phi_\alpha X_1^{\alpha_i}..X_n^{\alpha_n}, \psi_\beta Y_1^{\beta_i}..Y_m^{\beta_m}]$$

and expanding this expression using Lemma (3.1.2) shows that $[X,Y]\epsilon\Omega^o$, thus completing the proof.

□

§3.2: Examples

I) In this first 'example' we consider the possibilities of extending Th^m (3.1.4) to the case that the input noise process is a vector of dimension m and in so doing establish some connections with the following results of Ocone's.

THEOREM 3.2.1 (Ocone [2])

Consider the system

$$\Sigma \begin{cases} dx = f(x)dt + G\,dw \\ dy = h(x)dt + dv \end{cases}$$

with x evolving in some open connected set $V \subset \mathbb{R}^n$, $w \in \mathbb{R}^m$, G is a full rank matrix of dimension (n×m) with $m \geq n$ and $f \in \Gamma^\infty(V)$, $h \in C^\infty(V)$. Then, if $\dim\Lambda(\Sigma) < \infty$ and $\phi \in \Lambda(\Sigma) \cap C^\infty(V)$, ϕ is a polynomial of degree ≤ 2.

□

To begin our analysis we remark that in the case $m = n = 1$, the above result follows directly from our Th^m (3.1.4), for we now have that g(x) is a non-zero constant so

$$L_g^2(h) = g(x)\frac{\partial}{\partial x} g(x)\frac{\partial}{\partial x} h = g^2 \frac{\partial^2 h}{\partial x^2}$$

or

$$\frac{\partial^2 h}{\partial x^2} = \text{constant}$$

hence h is quadratic. For the case $m > 1$, however, there are some complications and it turns out that there are two, non-equivalent, generalisations we can consider. To see the problem let us consider the system

$$(3.2.1) \qquad \begin{cases} dx = f(x)dt + \sum_{i=1}^{m} g_i(x)dw^i & x \in \mathbb{R}^n \\ dy = h(x)dt + dv & y \in \mathbb{R}. \end{cases}$$

Now, as we pointed out earlier, in order to use the Pathwise concept of a solution to a stochastic differential equation we really need commutativity of the input vector fields. Thus our first constraint must be that

$$\mathcal{F} \triangleq \{g_1, \ldots, g_m\}_{L.A.}$$

is abelian. Next we need some sort of spanning hypothesis on $\mathcal{F}$ and this leads us to two possibilities. From Th^m (3.2.1) the natural assumption to make is that $\mathcal{F}$ is transitive on $\mathbb{R}^n$, whilst from Th^m (3.1.4) it is more obvious to assume that $\{g_1, \ldots, g_m\}$ are all linearly independent. The reasons for this second choice will become more obvious as we proceed, but before going further it should be pointed out that these criteria need not coincide; one need only consider the family of vector fields $\{x_i \frac{\partial}{\partial x_i};\ 1 \leq i \leq n\}$ to find an example of an n-dimensional, non-transitive abelian Lie Algebra.

We begin by assuming that F is transitive. Then, from Lemma (1.2.7), it follows immediately that around any point $x \in \mathbb{R}^n$ there is a coordinate chart (U,Φ) such that after a possible reordering

$$(\Phi_* g_k)(z) = \left.\frac{\partial}{\partial z_k}\right|_z \qquad \forall\ 1 \leq k \leq n,\ z \in \Phi(U)$$

(note: since F is transitive, $m \geq n$). Further, if we suppose that for $1 \leq j \leq m-n$

$$(\Phi_* g_{n+j})(z) = \sum_{i=1}^{n} \gamma_{ji}(z) \frac{\partial}{\partial z_i}$$

Then, $\forall k,j$

$$\begin{aligned}[\Phi_* g_k, \Phi_* g_{n+j}] &= [\frac{\partial}{\partial z_k}, \Sigma \gamma_{ji} \frac{\partial}{\partial z_i}] \\ &= \Sigma \frac{\partial \gamma_{ji}}{\partial z_k} \frac{\partial}{\partial z_i} \\ &= \Phi_*[g_k, g_{n+j}] \\ &= 0.\end{aligned}$$

Thus, $\frac{\partial \gamma_{ji}}{\partial z_k} = 0\ \forall\ 1 \leq k \leq n$ and so γ_{ji} are constant. Locally we can therefore transform (3.2.1) into the system

$$\Sigma_z \begin{cases} dz = \tilde{f}(z)dt + G\,dw \\ dy = \tilde{h}(z)dt + dv \end{cases}$$

and G is a matrix of full rank. Moreover, as shown in Brockett [3] and discussed in greater detail in §4.3, the diffeomorphism Φ induces a Lie algebra isomorphism between the estimation algebra Λ of (3.2.1) and $\Lambda(\Sigma_z)$. From Th^{m} (3.2.1) it therefore follows that if $\dim \Lambda < \infty$ $\tilde{h}$ must be quadratic, or in other words

$$\frac{\partial^2 h}{\partial z_i \partial z_j} = c_{ij} \qquad \forall\ 1 \leq i,m \leq n,\ \forall\ z \in \Phi(u)$$

for some constants c_{ij}. Now, for any $x \in U$, and $1 \leq i,j \leq n$

$$\begin{aligned}L_{g_i} L_{g_j}(h)(x) &= L_{\Phi_*^{-1} \frac{\partial}{\partial z_i}} L_{\Phi_*^{-1} \frac{\partial}{\partial z_j}}(h)(x) \\ &= L_{\frac{\partial}{\partial z_i}} L_{\frac{\partial}{\partial z_j}} (h \circ \Phi^{-1})\ (\Phi(x)) \\ &= \frac{\partial^2 \tilde{h}}{\partial z_i \partial z_j} (\Phi(x))\end{aligned}$$

But $\Phi(x) \varepsilon \Phi(U)$ so this last quantity is constant. It is trivial to see using the transitivity of F that in fact this identity is also valid for $1 \leq j, j \leq m$; so that we have shown that for any $x \varepsilon \mathbb{R}^n$ there is a neighbourhood U of x in $\mathbb{R}^n$ such that $L_{g_i} L_{g_j}(h)$ is constant on U for all i,j. Finally, this means that for all $1 \leq k \leq m$

$$L_{g_k}(L_{g_i} L_{g_i}(h))(x) = 0 \qquad \forall\ x \varepsilon \mathbb{R}^n$$

so from the Theorem of Abraham and Marsden quoted in §1.1 and the transitivity of F, $L_{g_i} L_{g_j}(h)$ is actually constant on the whole of $\mathbb{R}^n$.

Now let us turn our attention to the case that F is of dimension m. We first remark that in this situation the generator F of Λ becomes

$$F = - \Sigma \frac{\partial}{\partial x_i} f_i + \tfrac{1}{2} \Sigma \frac{\partial^2}{\partial x_i \partial x_j} (G(x) G^T(x))_{ij} - \tfrac{1}{2} h^2 ,$$

where G(x) is the matrix with columns $(g_1(x), \ldots, g_m(x))$, which in turn is given by

$$F = L^*_{\hat{f}} + \tfrac{1}{2} \sum_{i=1}^{m} L^{*2}_{g_i} - \tfrac{1}{2} h^2$$

and now $\hat{f} = f - \tfrac{1}{2} \Sigma \frac{dg_i}{dx} g_i$. In directly analagous fashion to Th^m (3.1.4) it is possible to show after tedious calculation that the principal part of $\text{ad}^k_F(G)$ is

$$\sum_{|\alpha|=k} L^{\alpha_1}_{g_1} \ldots L^{\alpha_m}_{g_m}(h) L^{*\alpha_1}_{g_1} \ldots L^{*\alpha_m}_{g_m}$$

However, from the Poincare-Birkhoff-Witt Theorem, the products $L^{*\alpha_1}_{g_1} \ldots L^{*\alpha_m}_{g_m}$ are all linearly independent so the constraint $\dim \Lambda < \infty$ now forces the existence of k s.t.

$$L^{\alpha_1}_{g_1} \ldots L^{\alpha_m}_{g_m}(h) \equiv 0 \qquad \forall\ |\alpha| \geq k \tag{3.2.2}$$

By again mimicing the proof of Th^m (3.1.4) we can construct a sequence of elements of Λ of order $(\bar{m}(\kappa-3) + 2)$ where $\kappa = \min\{k; (3.2.2)$ holds$\}$ with highest order term

$$\sum_{\substack{|\alpha| = \kappa-1 \\ |\beta| = \bar{m}(\kappa-3)+2}} (L^{\alpha_1}_{g_i} \ldots L^{\alpha_m}_{g_{n_1}}(h))^{\bar{m}} L^{*\beta_1}_{g_1} \ldots L^{*\beta_m}_{g_m}$$

from which we conclude, as before, that $L_{g_1}^{1} \ldots L_{g_m}^{m}(h)$ is constant $\forall |\alpha| = 2$.

We summarise the preceding discussions in the following Theorem.

THEOREM (3.2.2)

Consider the system (3 or 2.1) and assume that

(i) $\mathcal{F} = \{g_1, \ldots, g_m\}_{L.A.}$ is Abelian

(ii) either a) $\mathcal{F}$ is transitive on $\mathbb{R}^n$ or b) $\mathcal{F}$ has dimension m

(iii) the associated Estimation Algebra is finite dimensional.

Then, $L_{g_i} L_{g_j}(\phi)$ is constant, $\forall\ 1 \leq i,j \leq m$ and $\forall\ \phi \in \Lambda - 1$ □

As a final comment, we observe that Th^m (3.2.1) can be treated via the second method given above in the special case that m = n since we then find $g_1, \ldots, g_n$ (the columns of G) are all constant and are linearly independent by the rank condition on G. Now, if $[\alpha_{mn}] \overset{\Delta}{=} G^{-1}$ we have

$$\begin{aligned}
\frac{\partial^2 h}{\partial x_i \partial x_j} &= L_{\frac{\partial}{\partial x_i}} L_{\frac{\partial}{\partial x_j}} (h) \\
&= L_{\Sigma\alpha_{ik} g_k} L_{\Sigma\alpha_{j\ell} g_\ell}(h) \\
&= \Sigma\, \alpha_{ik}\alpha_{j\ell}\, L_{g_k} L_{g_\ell} (h) \\
&= \text{constant}
\end{aligned}$$

and so h is quadratic as required.

(II) As a further simple example, let us now consider the scalar system

$$(3.2.3)\qquad \begin{cases} dx_t = f(x)dt + p(x)dw_t \\ dy_t = q(x)dt + dv_t \end{cases}$$

with p a polynomial of degree n and q a polynomial of degree m, ie

$$p(x) = \sum_{p=0}^{n} p_k x^k \quad , \quad q(x) = \sum_{\ell=0}^{m} q_\ell x^\ell , \quad p_n q_m \neq 0$$

it is not clear that $m \leq 2$, despite the previous analysis, since we have not assumed $p(x) \neq 0 \ \forall x \in \mathbb{R}$. However, from Theorem (3.1.4), we know that $L_p^2(q)$ = const., and a simple calculation shows

$$L_p^2(q) = \sum_{\substack{\ell=0,..,m-1 \\ j,k=0,..,n}} (\ell+1)(k+\ell+1)p_j q_{\ell+1} p_k x^{j+k+\ell-1}$$

so

$$m(n+m)p_n^2 q_m x^{2n+m-2} = \text{constant}$$

We then have the following alternatives (since $m,n \in \mathbb{N}$ and $p_n q_m \neq 0$)

a) $m = 0$

b) $2(n-1) + m = 0$

i.e.

i) $q(x)$ is constant

ii) $m = -2(n-1)$

We have therefore shown that for the polynomial system (3.2.3) with q(x) non-constant the associated estimation algebra can only be finite dimensional if p(x) is constant and q is at most quadratic. This condition can also be shown to be sufficient under the further assumption that the drift vector field satisfies either the Benes or Generalised Benes hypotheses, namely

$$\text{a)}\quad \frac{df}{dx} + f^2 = ax^2 + bx + c$$

$$\left.\begin{array}{l} \text{b)}\quad \dfrac{df}{dx} + f^2 = -h^2 + a(2\alpha x + \beta)^2 + \dfrac{c}{(2\alpha x+\beta)^2} \\[2ex] \text{c)}\quad \dfrac{df}{dx} + f^2 = -h^2 + ax^2 + bx + c \end{array}\right\} \begin{array}{l} \text{if } h = \alpha x^2 + \beta x + \gamma \\ \alpha \neq 0. \end{array}$$

These systems were first considered in Benes important paper [1] which was amongst the first articles to rigourously demonstrate, using probabilistic techniques, the existence of finite dimensionally computable statistics for a nonlinear system. The Lie algebra analysis was studied in detail by Ocone [2] who further demonstrated that these are the only scalar systems for which the Wei-Norman construction is valid.

(III) In the previous section, we raised the point that even though the estimation algebra could satisfy Brocketts Homomorphism Principle, this did not give any guidance as to what class of statistics could be finite dimensionally computed. Hijab [2] has given a conceptual algorithm, based on Fliess' ideas on syntactic Lie algebras [3] , [4] which casts some light on this problem. Consider then the (possibly infinite dimensional) linear analytic system, $\Sigma = \{f,g,h\}$ with initial condition x_o, and define a map $\omega: \mathscr{L}(\Sigma)\to H_A(\Sigma)^*$, where $H_A(\Sigma)^*$ is the dual space of $H_A(\Sigma)$, by

$$\omega(X)(\phi) = L_X(\phi)(x_o)$$

The Macmillan degree, or rank, of Σ is then defined as the dim (Im ω). This dimension turns out to be an integer invariant of the input-output map.

THEOREM 3.2.3 (Fliess [3], Hijab [2])

The Macmillan degree of Σ is realisation invariant and is finite iff there is a finite dimensional system realising Σ.

□

In principle, then one can apply this result to the filter (3.1.8) to investigate the existence of a representation of the form (3.1.9). The complexity of the previous analysis suggests that this may not be an easy task. However, Hijab used it with some degree of success in analysing the system

$$(3.2.4) \quad \begin{cases} \dot{x} = f(x) & x(0) = x_o \ , \ x \in \mathbb{R}^n \\ dy = h(x)dt + dv_t & \end{cases}$$

where x_0 is a random variable with full support (ie if $\phi : \mathbb{R}^n \xrightarrow{} \mathbb{R}^+$ s.t. $E(\phi(x_o)) = 0$ then $\phi \equiv 0$), and showed that the output was f.d.c. iff $Sp\{L_f^k(h);\ k \geqslant 0\}$ was finite dimensional. As a further n.a.s.c., this latter criterion is equivalent to requiring that (3.2.4) be a nonlinear realisation of a linear system

$$\begin{cases} \dot{z} = Az & z(0) = z_o \ , \quad z \in \mathbb{R}^N \\ dy = Cz\, dt + dv_t & \end{cases}$$

Moreover it is easily seen that for (3.2.4)

$$\Lambda(\Sigma) = Sp\{L_f^* - \tfrac{1}{2}h^2, L_f^k(h);\ k \geqslant 0\}$$

and that the identification

$$L_f^* \to \Sigma f_i(x)\frac{\partial}{\partial x_i} \ , \ \phi(x) \to \phi(x)\frac{\partial}{\partial \xi} \qquad \forall \phi \in \Lambda_{-1}$$

provides a Lie algebra homomorphism between $\Lambda(\Sigma)$ and $\Gamma^\omega(T\mathbb{R}^{n+1})$. Thus, Brocketts principle is always satisfied, but $\hat{h}$ can only be computed in a finite dimensional way if $\Lambda(\Sigma)$ is finite dimensional. A recent paper of Levine, [1], provides a rigourous probabilistic proof of this result.

Using much the same techniques, the system below also proves amenable to this analysis

$$\Sigma_u \quad \begin{cases} \dot{x} = f(x) + \sum_{i=1}^{m} u_i(t)g_i(x) & x(0) = x_o \ x \in \mathbb{R}^n \\ dy = h(x)dt + dv & \end{cases}$$

Here the inputs, u_i, are taken to be deterministic control functions which are allowed to be piecewise constant and, as before, x_o is a random variable of full support with density ρ_o (we should remark that both Σ_u and (3.2.4) are supposed to represent the situation that noisy observations are taken of a deterministic system with random initial condition). In this case the generators are given by $\{L_f^* - \frac{1}{2}h^2, L_{g_i}^*, h; 1 \leq i \leq m\}$. (As in the purely deterministic case, one switches the controls on and off arbitrarily to decompose the single generator $L_f^* + \Sigma_{u_i} L_{g_i}^* - \frac{1}{2}h^2$ into the above components). It is then easy to see that

$$\Lambda(\Sigma_u) = \mathcal{H}(\Sigma_u) \oplus \{L_f^* - \tfrac{1}{2}h^2, L_{g_i}^*\}_{L.A.} \tag{3.2.5}$$

$$\subset \mathcal{H}(\Sigma_u) \oplus [\{L_f^*, L_{g_i}^*\}_{L.A.} + \mathcal{H}(\Sigma_u) \oplus \mathcal{H}(\Sigma_u)]$$

Again, it is immediate that $\Lambda(\Sigma_u)$ satisfies Brocketts principle under the homomorphism $(L_X^* + \phi) \to L_X + \phi\frac{\partial}{\partial\xi}$ for some 'dummy' variable ξ s.t. $\frac{\partial}{\partial\xi}$ commutes with ϕ and L_X. Suppose, now, that we wish to compute $\hat{h}(t) = C_h(\rho)(t)$ (where $C_h(\rho)$ is as defined in (3.1.7)). Then in order to apply Hijab's algorithm we must first calculate the observation space (resp. algebra) of the filter which we denote by $\mathcal{H}^\Lambda$(resp. $\mathcal{H}_A^\Lambda(\Sigma_u)$). Now, for $X \in \Lambda$, with trajectory γ_t^X defined on some neighbourhood of $t = 0$, we readily obtain (using the chain rule)

$$L_X(C_h)(\rho_o) \triangleq \frac{d}{dt} C_h(\gamma_t^X \rho_o)\Big|_{t=0}$$

$$= \int h(x)X(\rho_o)dx \left(\int \rho_o dx\right)^{-1} - \frac{\int h\rho_o dx \int X(\rho_o)dx}{(\int \rho_o dx)^2}$$

$$= C_{X^*(h)}(\rho_o) - C_h(\rho_o)C_{X^*(1)}(\rho_o) \tag{3.2.5}$$

with X^* denoting the (formal) adjoint of X. Now let us define the following space of functions

$$\mathbb{P} = \mathrm{Sp}\{C_{\phi_1} \ldots C_{\phi_n} ; n \geq 1, \phi_j \in \mathcal{H}_A(\Sigma_u)\}$$

Clearly, from (3.2.6), $\mathbb{P}$ is invariant under the action of $\Lambda(\Sigma_u)$ and contains C_h. Thus $\overset{\Lambda}{\mathscr{H}}_A \subset \mathbb{P}$. Conversely we first see that from (3.2.5) $\mathscr{H}(\Sigma_u) \subset \Lambda(\Sigma_u)$ so $\{L_\phi(C_h); \phi \varepsilon \mathscr{H}(\Sigma_u)\} \subset \overset{\Lambda}{\mathscr{H}}(\Sigma_u)$. But, by (3.2.6), if $C_\psi \varepsilon \overset{\Lambda}{\mathscr{H}}_A(\Sigma_u)$, then

$$L_\phi(C_\psi) = C_{\phi\psi} - C_\phi C_\psi. \tag{3.2.7}$$

from which we deduce, by setting $\phi = h$, and ψ equal first to h, then h^2, that both C_{h^2} and C_{h^3} $\varepsilon \overset{\Lambda}{\mathscr{H}}_A(\Sigma_u)$. However, $\overset{\Lambda}{\mathscr{H}}_A(\Sigma_u)$ must be invariant under the action of $L_f^* - \tfrac{1}{2}h^2$ and $L_{g_i}^*$. In particular, this means

$$L_{(L_f - \frac{1}{2}h^2)}(C_h) = C_{L_f(h) - \frac{1}{2}h^3} + C_h\, C_{h^2/2} \quad \varepsilon\, \overset{\Lambda}{\mathscr{H}}(\Sigma_u)$$

$$L_{L_{g_i}^*}(C_h) = C_{L_{g_i}(h)} \qquad 1 \leqslant i \leqslant m$$

so that $C_{L_f(h)}$, $C_{L_{g_i}(h)}$ $\varepsilon \overset{\Lambda}{\mathscr{H}}_A(\Sigma_u)$. Inductively, it is not difficult to see that this implies $\{C_\phi; \phi \varepsilon \mathscr{H}(\Sigma_u)\} \subset \overset{\Lambda}{\mathscr{H}}_A(\Sigma_u)$ and iterative use of (3.2.7) shows then that $\mathbb{P} \subset \overset{\Lambda}{\mathscr{H}}_A(\Sigma_u)$.

We now claim that $\dim \mathscr{H}(\Sigma_u) \leqslant \text{rank}\ {}^{\Lambda}(\Sigma_u) + 1$. First note that if $c \varepsilon \mathbb{R}$ then from (3.2.7)

$$L_c(C_h) \equiv 0$$

$$\begin{aligned} \text{Thus, rank}\ {}^{\Lambda}(\Sigma_u) &= \dim \omega(\{\Lambda(\Sigma_u), \mathbb{R}\}_{L.A.}) \\ &= \dim \omega\{(\mathscr{H}(\Sigma_u) + \mathbb{R}) \oplus \{L_f^* - \tfrac{1}{2}h^2, L_{g_i}^*\}_{L.A.}\} \end{aligned}$$

so $\dim \omega(\mathscr{H}(\Sigma_u) + \mathbb{R} \leqslant \text{rank}\ {}^{\Lambda}(\Sigma_u)$. Now suppose that $\phi \varepsilon \mathscr{H}(\Sigma_u) + \mathbb{R}$ satisfies $\omega(\phi) = 0$. Then $\forall \Phi \varepsilon \{\mathscr{H}_A(\Sigma_u) + \mathbb{R}\}$

$$\begin{aligned} \omega(\phi)(C_\Phi) &= L_\phi(C_\Phi)(\rho_o) \\ &= (C_{\phi\Phi} - C_\phi C_\Phi)(\rho_o) \\ &= C_{(\phi - C_\phi(\rho_o))\Phi}(\rho_o) \end{aligned}$$

But $C_\phi(\rho_o)$ is constant, so we can choose $\Phi = \phi - C_\phi(\rho_o)$ to obtain

$$C_{(\phi - C_{\phi}(\rho_o))^2}(\rho_o) = 0$$

and ρ_o is of full support. Hence, $\phi = C_{\phi}(\rho_o) \varepsilon \mathbb{R}$ and so, finally, we see

$$\dim \ \omega(\mathcal{H}(\Sigma_u) + \mathbb{R}) = \dim \mathcal{H}(\Sigma_u) - 1 \leq \ \text{rank } \Lambda(\Sigma_u)$$

When we couple this result with Th^m (3.2.3), we see that if $\hat{h}$ is f.d.c. then rank $^{\Lambda}(\Sigma_u)$ and hence $\dim \mathcal{H}(\Sigma_u)$ is finite. An immediate corollory of this is that the underlying deterministic system of this problem has a bilinear realisation: simply choose a basis $\{\phi_1, \ldots, \phi_N\}$ for $\mathcal{H}(\Sigma_u)$ and define $z(t) = (\phi_1(x(t)), \ldots, \phi_N(x(t)))^T$. Then

$$\begin{aligned} \dot{z}_j &\overset{\Delta}{=} \frac{d}{dt}\, \phi_j(x(t)) = L_f(\phi_j)(x(t) + \Sigma u_i(t) L_{g_i}(\phi_j)(x(t)) \\ &= \sum_{k=1}^{N} \alpha_{ik} \phi_k(x(t)) + \Sigma u_i(t) \beta^i_{jk} \phi_k(x(t)) \end{aligned}$$

since $L_f(\phi_j), L_{g_i}(\phi_j) \varepsilon \mathcal{H}(\Sigma)$. Thus, z satisfies,

$$\dot{z} = (A + \Sigma u_i B_i) z(t) \qquad z_j(0) = \phi_j(x(0)) \ \ 1 \leq j \leq n$$

and

$$\begin{aligned} y(t) = h(x(t)) &= \Sigma c_j \phi_j(x(t)) \\ &= Cz(t). \end{aligned}$$

We summarise the preceding discussion in the following Theorem.

<u>THEOREM 3.2.4</u>

Consider the minimal system

$$\Sigma_u \left\{ \begin{array}{ll} \dot{x} = f(x) + \sum_{i=1}^{m} u_i g_i(x) & x \varepsilon \mathbb{R}^n, \ x(0) = x_o \\ & \\ dy = h(x)dt + dv & y \varepsilon \mathbb{R} \end{array} \right.$$

with all hypothesis as above. Then for the following statements

(i) $\hat{h}$ is f.d.c. (ii) Σ_u has a bilinear realisation

(iii) $\Lambda(\Sigma_u)$ is finite dimensional

we have (i) => (ii) <=> (iii).

Proof

(i) $\Rightarrow$ (ii) is exactly the argument given above. To see that (ii) is equivalent to (iii) we need only note identity (3.2.5). For if $\dim \Lambda(\Sigma_u) < \infty$ then $\dim \mathscr{H}(\Sigma_u) < \infty$ and the above analysis applies directly. Whilst, if Σ_u satisfies (ii), then both $\mathscr{H}(\Sigma_u)$ and $\mathscr{L}(\Sigma_u)$ are finite dimensional and, hence, so is $\Lambda(\Sigma_u)$.

□

Before leaving this example, we take the opportunity to make some remarks concerning the added complexities of the algebraic estimation problem which arise if the estimation algebra is no longer assumed to be finite dimensional; indeed, from the restrictive nature of the conditions derived in all the previous analyses, it seems that this hypothesis will seldom be satisfied. The first comment we pass extends this argument slightly. As a guiding principle, the initial step usually made in generalising finite dimensional analysis to more abstract spaces is to assume that there is still some Banach structure to draw on.

However, in the Lie algebra sense (or even in the theory of more general Banach algebras) it is usually assumed that the operation of taking the product is also continuous with respect to this topology. In particular, this means that if $\mathscr{L}$ is a Banach Lie algebra (B.L.A.) then the underlying vector space has a complete, normed topology s.t. $\forall X \in \mathscr{L}, \mathrm{ad}_X : \mathscr{L} \to \mathscr{L}$ is a bounded linear operator. This simple fact allows for the immediate construction of counter examples to the conjecture that the estimation algebra is Banach. For, suppose that we wish to calculate $\Lambda(\Sigma_u)$ (with Σ_u as in Th^{m} (3.2.4) for the specific case that

$$f(x) = x\frac{\partial}{\partial x}\ , \ g_1(x) = x^2\frac{\partial}{\partial x}\ , \ g_2(x) = x^3\frac{\partial}{\partial x} \text{ and } h(x) = x$$

then the generators are given by

$$\{\ \frac{\partial}{\partial x}x + \tfrac{1}{2}x^2,\ \frac{\partial}{\partial x}x^2,\ \frac{\partial}{\partial x}x^3,\ x\ \}$$

A trivial application of Lemma (3.1.3) immediately shows

$$[\frac{\partial}{\partial x}x^2, x] = -x^2$$

and, hence, $\frac{\partial}{\partial x} x \in \Lambda(\Sigma_u)$. Moreover, it is readily seen that (modulo a constant, non zero factor)

$$\mathrm{ad}^k_{\frac{\partial}{\partial x}x^2} \left(\frac{\partial}{\partial x} x^3\right) = \frac{\partial}{\partial x} x^{k+3} \qquad k \geq 0$$

Now let us assume that $\Lambda(\Sigma_u)$ is a B.L.A. In particular, this implies that $\mathrm{ad}_{\frac{\partial}{\partial x}x} : \Lambda \to \Lambda$ is bounded. But, $\forall\ k > 1$, as we have seen already, $\frac{\partial}{\partial x} x^k \in \Lambda(\Sigma_u)$, and

$$\mathrm{ad}_{\frac{\partial}{\partial x}x} \left(\frac{\partial}{\partial x}x^k\right) = (k-1) \frac{\partial}{\partial x} x^k$$

Taking norms we then find

$$(k-1)\left|\left| \frac{\partial}{\partial x}x^k \right|\right| \leq \left|\left| \mathrm{ad}\, \frac{\partial}{\partial x}x \right|\right| \, \left|\left| \frac{\partial}{\partial x}x^k \right|\right| \qquad \forall k > 1.$$

contradicting the boundedness of $\mathrm{ad}_{\frac{\partial}{\partial x}x}$. Thus, Λ cannot have a Banach structure.

[As an aside, note that this example also illustrates the insufficiency of Brocketts Principle, for, as we have already remarked, $\Lambda(\Sigma_u)$ is isomorphic to a Lie algebra of vector fields on $\mathbb{R}^2$. But from Th^m (3.2.4), $\hat{x} = \hat{h}$ is not f.d.c. since the corresponding observation space is infinite dimensional and, in fact, contains $\mathbb{R}[x]\backslash\mathbb{R}$].

The set up described above has analogies with the problem considered in Omori [1] and Omori, de la Harpe [1], of classifying those Banach Lie groups acting smoothly on a finite dimensional manifold. Let us assume that $\hat{\psi}$ is an f.d.c. statistic and that the corresponding estimation algebra Λ is Banach. We denote by π the Brockett homomorphism taking Λ into the Lie algebra of vector fields on the state manifold, M, of $\hat{\psi}$.

If we further suppose that π is also continuous with respect to the usual topology on $\Gamma^{\infty}(TM)$, then the image $\pi(\Lambda)$ can clearly be given a Banach structure. As we remarked previously, without loss of generality we can take the realisation of $\dot{\psi}$ on M to be minimal, and if we also assume that $\pi(\Lambda)$ is a Lie algebra of complete vector fields, then, by Theorem A of Omori [1], there is a Banach Lie subgroup G of Diff (M) which (by minimality) acts smoothly, effectively and transitively on M. This imposes immediate restrictions on $\pi(\Lambda)$ as the following result demonstrates.

THEOREM 3.2.5 (Omori [1])

Let G be a connected Banach Lie group acting smoothly, effectively and transitively on a (finite dimensional) manifold M. Then

a) if M is compact, G is finite dimensional

b) if M is non-compact, G is almost solvable, ie the Lie algebra $\mathfrak{g}$ of G contains a solvable, finite codimensional, closed ideal ρ (solvability in this case requires that if $p_o = p$, and p_n is defined as the closure of $[p_{n-1}, p_{n-1}]$ then $\exists\ N<\infty$ s.t. $p_{N+1} = \{0\}$).

□

Of course, in the case that the estimation algebra is finite dimensional, Th^{m} (3.2.5b) is an immediate consequence of Levis Theorem that any finite dimensional Lie algebra is the direct sum of a solvable ideal with a semi-simple subalgebra (Jacobson [2]). The full implications of Th^{m} (3.2.5) in the present context have yet to be explored, but Banach Lie groups have been generated by considering parameter estimation algorithms as nonlinear filtering problems (Krishnaprasad, Hazewinkel and Marcus [1], [2], [3]). However, from the above remarks it seems clear that, in general, some weaker topology on the estimation algebra will be found. In some sense, this brings us full circle, since Fliess' construction of the MacMillan degree is based in turn on the work of

Singer and Sternberg [1] and Guillemin and Sternberg [1], who show that any linearly compact Lie algebra possessing a fundamental subalgebra is isomorphic to a Lie algebra of formal vector fields on a finite dimensional vector space. Without going into too much details, for which we refer to the recent text of Conn [1], we remark that a subalgebra L_o of a complete topological Lie algebra L is fundamental if it has finite codimension and the induced chain of subalgebras

$$L_i = \{X \varepsilon L_{i-1} \; ; \; [X,Y] \; \varepsilon \; L_{i-1} \; \forall Y \varepsilon L\}$$

forms a fundamental system of neighbourhoods of the origin and $\bigcap_{i\geq 0} L_i = \{0\}$. Thus, the topology on L is much weaker than that induced by a norm, however Th^m(3.2.5(b)) does have a parallel (Conn [1], Th^m 1.1) since L also satisfies a descending chain condition on closed ideals. Other connections can be made and this is clearly an area which could be usefully further researched.

V) We close this section, and the chapter, by remarking that the necessary condition derived in Th^m(3.1.4) is trivially satisfied by the class of systems, studied originally by Marcus and Willsky [1], taking the form

$$(3.2.8) \qquad \left\{ \begin{array}{ll} \left\{ \begin{array}{l} dx^1 = Ax^1 dt + B\,dw \\ dx^2 = f(x^2)dt + G(x^2)dx^1_t \end{array} \right. & x^i \varepsilon \mathbb{R}^{n_i} \qquad i = 1,2 \\ \\ \quad dy = Cx^1 dt + dv & \end{array} \right.$$

where $\dot{x}^2 = f(x^2)dt + \sum_{j=1}^{n_1} G_j(x^2)u_j$ has a finite Volterra series. For such systems it can be shown that statistics of the x^2 process which are f.d.c. do exist - thus they form one of the few known such classes exhibiting truely nonlinear behaviour. It also turns out that the associated estimation algebras have a strong algebraic structure and possess many ideals. This structure has been fully explored in

Hazewinkel, Liu and Marcus [1] and related papers. (We cannot leave this example without pointing out the obvious: Linear systems are included in the class of systems defined by (3.2.8). In this case, the calculation of the estimation algebra is quite straightforward and it turns out to be both solvable and finite dimensional (Brockett [2])).

Hazewinkel, Liu and Marcus [1] and related papers. (We cannot leave this example without pointing out the obvious: Linear systems are included in the class of systems defined by (3.2.8). In this case, the calculation of the estimation algebra is quite straightforward and it turns out to be both solvable and finite dimensional (Brockett [2])).

CHAPTER IV: AN EXAMPLE OF HAZWINKEL-MARCUS

In this, the final chapter of the thesis, we synthesise various ideas developed in the previous chapters in order to investigate several points raised by the following example

$$(4.0.1)\qquad \begin{cases} dx_1 = dw_1 \\ dx_2 = x_1 dt + x_1 dw_2 \\ dy = x_2 dt + dv \end{cases}$$

The underlying deterministic structure of this system is that of a graded polynomial form on $\mathbb{R}^2$, with gradation $\mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2}, n_1 = n_2 = 1$, so it has associated with it the algebraic properties of such systems as described in Chapters I and II. Moreover, the input vector fields $g_1 = \frac{\partial}{\partial x_1}$ & $g_2 = x_1 \frac{\partial}{\partial x_2}$ are linearly independent and satisfy

$$L_{g_i} L_{g_j}(h) = \text{constant} \in \{0,1\}$$

where h is the output function $h(x) = x_2$, so appearing to comply with the necessary condition for finite dimensionality of the estimation algebra, except, of course, $\{g_1, g_2\}_{L.A.}$ is not abelian. It might therefore be expected that the filtering properties of this system should be 'nice'. This indeed turns out to be the case, but not in the positive sense to be desired.

The problem is that, as shown in Hazewinkel-Marcus [1] (where the example was first studied from this point of view), the estimation algebra of (4.0.1) is W_2, the Weyl algebra on 2-generators, where in general we shall assume that W_n is the faithful representation of the abstract Weyl algebra on n-generators given by

$$W_n = \{\sum_{|\alpha|=0}^{k} \phi_\alpha \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}} \; ; \; k \geq 0, \phi_\alpha \in \mathbb{R}[x_1, \ldots, x_n]\}$$

Thus, for our purposes, W_n is the Lie algebra of differential operators on $\mathbb{R}^n$ with polynomial coefficients. For algebraic estimation the significance of this calculation lies in the following result.

THEOREM 4.0.1 (Hazewinkel and Marcus [1])

(i) As a Lie algebra, W_n is generated by

$$\{x_i, \frac{\partial^2}{\partial x_i^2}, x_i^2 \frac{\partial}{\partial x_i}, x_j x_{j+1}; \; i = 1,..,n, \; j = 1,...,n-1\}.$$

(ii) There are no nontrivial homomorphisms from W_n into either $\Gamma^\omega(TM)$ or $\Gamma^\infty(TM)$ for any finite dimensional manifold M.

$\square$

As an immediate corollary of Brocketts Principle we therefore deduce that there are no nontrivial f.d.c. statistics of any process whose estimation algebra is isomorphic to W_n. In particular, this observation applies to (4.0.1) so it is in the sense of non-existence that, despite the rich algebraic structure already established, the process has 'nice' filtering properties. This result is therefore quite surprising, not only for the reasons already described but also because (4.0.1) is one of the simplest of nonlinear systems. It is natural to ask then, how general this behaviour is and, in the sequel, by limiting attention to g.p. forms we go some way towards answering this point with the construction of a class of systems having estimation algebras isomorphic to W_n.

First, though, in §4.1 the estimation algebra for an arbitrary minimal system Σ in g.p.f. on $\mathbb{R}^n$ is studied particularly with regard to the general containment condition derived in Th^m (3.1.6). Due to the polynomialnature of Σ it is obvious that $\Lambda(\Sigma) \subset \Omega(\Sigma) \subset W_n$ and using the results of Chapter I it is not difficult to show that if $\hat{\Sigma}$ is also in minimal g.p.f. then $\Omega(\Sigma) = \Omega^o(\hat{\Sigma}\ ) = W_n$. (Recall that $\hat{\Sigma}$ is the system obtained from $\Sigma = \{f,g,h\}$ by the perturbation $f \to \hat{f} = f - \frac{1}{2}\frac{dg}{dx}g$, and $\Omega(\Sigma), \Omega^o(\hat{\Sigma})$ are tensor spaces). However, whilst the graded structure of Σ can be readily shown to be preserved, minimality of $\hat{\Sigma}$ is not guaranteed in general. Conditions are derived in Th^m (4.1.1) under which this will be the case.

As the next step in our construction of the required class of systems, we adapt a strong observability concept, introduced by Gauthier and Bornard [1] to develop a canonical form for certain single input-single output systems. The representation thus obtained is seen to closely resemble the structure of (4.0.1) except, of course, (4.0.1) has two input channels. However it can be shown through direct and tedious calculations that the system

$$\begin{cases} dx_1 = dw \\ dx_2 = x_1 dt + x_1 dw \end{cases}$$
$$dy = x_2 dt + dv$$

still has $\Lambda = W_2$, so (4.0.1) remains our 'inspiration'. The full computations required to show this are omitted as they form the basis for the analysis of §4.3 in which we finally obtain our class of systems satisfying $\Lambda \cong W_n$. The results obtained are still unsatisfactory since we have to assume that certain generators have already been established as elements of the estimation algebra, and further work is required to weaken these hypotheses. On the positive side, however, our theorem only requires that 3 elements be found compared with the (4n−2) of Th^m (4.0.1(i)).

§4.1 Graded Polynomial Forms, Algebraic Estimation and W_n

At the end of §3.2 we proved a general containment condition placing the estimation algebra Λ of an arbitrary linear analytic system $\Sigma = \{f,g,h\}$ within a tensor space related to the observation and strong accessibility algebras of the system $\hat{\Sigma} = \{\hat{f} \triangleq f - \frac{1}{2}\frac{dg}{dx}\,g, g, h\}$. In fact, be defining $\Omega^{o}(\hat{\Sigma}) = \mathcal{H}_{A}(\hat{\Sigma}) \otimes \mathcal{U}(\mathcal{S}(\hat{\Sigma}^{*}))$ we showed that

$$(4.1.1) \qquad \Lambda \subset \mathbb{R}F + \Omega^{o}(\hat{\Sigma}) = \mathbb{R}F + \mathcal{H}_{A}(\hat{\Sigma}) \otimes \bigoplus_{j \geq 0} \mathcal{S}(\Sigma^{*})^{\otimes j}$$

where, by convention $\mathcal{S}^{\otimes o} = \mathbb{R}$ and $\mathcal{S}^{\otimes j} = \mathcal{S}\otimes \ldots \otimes\mathcal{S}$, j-factors. Let us now assume that $\hat{\Sigma}$ is actually a minimal system in g.p.f. on $\mathbb{R}^{n}$. Then, since $\mathcal{S}(\hat{\Sigma}) \subset V_{1}$ and, as we saw in Th^{m} (1.1.2), with respect to the state space gradation $\bigoplus_{\ell=1}^{p} \mathbb{R}^{n_{\ell}}$, $V_{1} = \bigoplus_{\ell=1}^{p} Q^{\ell-1} \otimes \Delta_{\ell}$ it follows immediately that every $X \varepsilon \mathcal{S}(\hat{\Sigma})$ is skew adjoint so $\mathcal{S}(\hat{\Sigma}^{*}) = \mathcal{S}(\hat{\Sigma})$. The algebraic structure theorems proved in Chapters I and II now imply that

$$\begin{aligned}\Omega^{o}(\hat{\Sigma}) &= \bigoplus_{j \geq 0} \mathcal{H}_{A}(\hat{\Sigma}) \otimes \mathcal{S}(\hat{\Sigma})^{j} \\ &= \bigoplus_{j \geq 0} \mathbb{R}[x_{1},\ldots,x_{n}] \otimes \mathcal{S}(\hat{\Sigma})^{\otimes j} \\ &= \bigoplus_{j \geq 0} \tilde{D}_{j}\,(\mathbb{R}^{n})\end{aligned}$$

and so

$$(4.1.2) \quad \Omega^{o}(\hat{\Sigma}) = W_{n}.$$

Moreover, under these conditions, F is readily seen to be a differential operator with polynomial coefficients and hence is actually an element of $\Omega^{o}(\hat{\Sigma})$. In fact, we show later that $\hat{\Sigma}$ is in g.p.f. so $\hat{f} \varepsilon V_{o}$. Then for any smooth function ϕ

$$\begin{aligned}L_{\hat{f}}^{*}(\phi) &= -\sum \frac{\partial}{\partial x_{i}} \hat{f}_{i}\phi = -\sum \hat{f}_{i} \frac{\partial \phi}{\partial x_{i}} - \phi \frac{\partial \hat{f}_{i}}{\partial x_{i}} \\ &= -L_{\hat{f}}(\phi) - (\operatorname{div} \hat{f})\phi\end{aligned}$$

and from the graded structure we know that $\hat{f}_{i}$ is at most linear in the coordinate x_{i}. Thus, div $\hat{f}$ is constant so in this particular case

$$\begin{aligned}F &= L_{\hat{f}}^{*} + \tfrac{1}{2}L_{g}^{*2} - \tfrac{1}{2}h^{2} \\ &= -L_{\hat{f}} + \tfrac{1}{2}L_{g}^{2} - \tfrac{1}{2}h^{2} - c\end{aligned}$$

clearly demonstrating the polynomial nature of F. Equation (4.1.1) can therefore be reduced to

$$\Lambda \subset \Omega^0(\hat{\Sigma}) = W_n$$

In some sense this is not surprising since both generators of Λ will be elements of W_n and consequently it is obvious that $\Lambda \subset W_n$. What we have actually achieved here is the demonstration that $\Omega^0(\hat{\Sigma}) = W_n$ which is clearly non-trivial, and also shortens the (trivially proved) chain $\Lambda \subset \Omega^0(\hat{\Sigma}) \subset W_n$. For the remainder of this section we intend to investigate how true equation (4.1.2) remains if we only assume Σ is minimal and in g.p.f. In other words, we are asking the question how does the Itô correction term, $\frac{1}{2} \frac{dg}{dx} g$, affect the structure of the system Σ?. Ideally, of course, we should like to be able to show that Σ is minimal and in g.p.f. <=> $\hat{\Sigma}$ is in g.p.f. and minimal, but whilst the graded structure can be shown to carry over quite readily, minimality does present some problems.

The first point to notice is that since Σ is assumed to be in g.p.f. then $\hat{\Sigma}$ is in g.p.f. with respect to the same gradation of the state space as that of Σ. (In particular, from the results of §2.2, both Σ and $\hat{\Sigma}$ are realisations of stationary finite Volterra series although these input-output maps may be different). The initial claim that $\hat{\Sigma}$ is in g.p.f. is quite easy to prove. Indeed, if we denote by g' the vector field determined by $\frac{dg}{dx} g$ it is readily seen that the i^{th} component of g' is

$$g_i' \triangleq \sum_{j=1}^{n} g_j(x) \left. \frac{\partial g_i}{\partial x_j} \right|_x = L_g(g_i)(x)$$

However, by definition of the g.p.f., $g \varepsilon V_1$ and using the decomposition of V_1 given above, we get

$$\begin{aligned} g' &\varepsilon \bigoplus_{j=1}^{p} L_g(Q^{j-1}) \otimes \Delta_j \\ &\subset \bigoplus Q^{j-2} \otimes \Delta_j \triangleq V_2 \end{aligned}$$

and, hence,

$$\hat{f} \triangleq f - \tfrac{1}{2}\frac{dg}{dx}g \in V_o + V_2$$

$$\subset V_o$$

Thus, $\hat{\Sigma}$ is a linear analytic system defined on the graded space $\oplus\, \mathbb{R}^{n_j}$ and satisfying $\hat{f}\in V_o$, $g\in V_1$ and $h\in Q^r$, this last fact also following from the assumption that Σ is in g.p.f. In other words, $\hat{\Sigma}$ is in g.p.f.

Let us now turn our attention to the problem of determining when minimality of Σ implies minimality of $\hat{\Sigma}$. In general, this will not be true as the system

$$\begin{cases} \dot{x}_1 = u \\ \dot{x}_2 = x_1 - x_1^2 u \end{cases}$$

defined on the graded space $\mathbb{R}^2 = \mathbb{R} \oplus \mathbb{R}^0 \oplus \mathbb{R}$ shows. For this example we find that

$$f = x_1 \frac{\partial}{\partial x_2}, \quad g = \frac{\partial}{\partial x_1} - x_1^2 \frac{\partial}{\partial x_2}, \quad [f,g] = -\frac{\partial}{\partial x_2}$$

and all other brackets are zero so the system is strongly accessible. However,

$$\frac{dg}{dx}g = \begin{bmatrix} 0 & 0 \\ -2x_1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -x_1^2 \end{bmatrix} = \begin{bmatrix} 0 \\ -2x_1 \end{bmatrix}$$

Thus $\hat{f} \equiv 0$ which means $\hat{\Sigma}$ cannot be strongly accessible ($\mathscr{S}(\hat{\Sigma}) = \mathrm{Sp}\{g\}$ and so fails to be transitive). As the following result shows, the difficulty arises largely because the system is not in g.c.p.f.

<u>THEOREM 4.1.1</u>

Let Σ and $\hat{\Sigma}$ be the two linear analytic systems on $\mathbb{R}^n$ described above. Then

(i) Σ is in g.c.p.f. $\Leftrightarrow$ $\hat{\Sigma}$ is in g.c.p.f.

(ii) Σ satisfies the g.o.p.f. rank condition $\Leftrightarrow$ $\hat{\Sigma}$ satisfies the g.o.p.f. rank condition.

In particular, if Σ is in symmetric p.f. (ie, in both g.o. and g.c.p.f.) then $\hat{\Sigma}$ is minimal and in symmetric p.f. [Note also that (ii) does not mean that Σ and $\hat{\Sigma}$ are observable since we have not assumed accessibility].

Proof

First we note that Σ and $\hat{\Sigma}$ only differ by the addition or subtraction of the term $\frac{1}{2}\frac{dg}{dx}g$ to $\hat{f}$ or f respectively. It therefore suffices to prove the implications in only one direction. We intend to show that

$$\text{(i)} \quad \mathcal{S}^k(\Sigma)(0) = \bigoplus_{j=k}^{p} \mathbb{R}^{n_j} \Rightarrow \mathcal{S}^k(\hat{\Sigma})(0) = \bigoplus_{j=k}^{p} \mathbb{R}^{n_j}$$

and

$$\text{(ii)} \quad d\tilde{H}^k(\Sigma)(0) = \bigoplus_{j=1}^{\min(q,r-k)} \mathbb{R}^{m_j} \Rightarrow d\tilde{H}^k(\hat{\Sigma})(0) = \bigoplus_{j=1}^{\min(q,r-k)} \mathbb{R}^{m_j}$$

where we have assumed that for the g.c.p.f. (resp. the g.o.p.f.) the gradation on $\mathbb{R}^n$ is given by $\bigoplus_{j=1}^{p} \mathbb{R}^{n_j}$ (resp $\bigoplus_{j=1}^{q} \mathbb{R}^{m_j}$) and that the output function $h \varepsilon Q^r$.

Consider first the sequence of subspaces of $\mathcal{S}(\Sigma)$ defined by

$$R^1(\Sigma) = \mathrm{Sp}\{\mathrm{ad}_f^{\ell} g \; ; \; \ell \geq 0\}$$

$$R^{k+1}(\Sigma) = [R^1(\Sigma), R^k(\Sigma)]$$

(with appropriate adaptation for $\hat{\Sigma}$). Then we claim that

$$\text{(4.1.3)} \qquad R^k(\hat{\Sigma}) = R^k(\Sigma) \bmod V_{k+1} \qquad 1 \leq k \leq p$$

To prove this, suppose that $k=1$. Then, clearly, $g \,\varepsilon\, R^1(\Sigma) \cap R\ (\hat{\Sigma})$. Assume, inductively, that $\mathrm{ad}_{\hat{f}}^{\ell}(g)$ can be written as a sum X_1+X_2 with $X_1 \varepsilon R^1(\Sigma)$ and $X_2 \varepsilon V_2$ for $\ell = 0\ldots,L$. Then

$$\begin{aligned} \mathrm{ad}_{\hat{f}}^{L+1}(g) &= \mathrm{ad}_{\hat{f}}\ (X_1+X_2) = \mathrm{ad}_{f-\frac{1}{2}g'}(X_1+X_2) \\ &= \mathrm{ad}_f(X_1) + \mathrm{ad}_f(X_2) - \tfrac{1}{2}\{\mathrm{ad}_{g'}(X_1) + \mathrm{ad}_g(X_2)\} \\ &\varepsilon\ R^1(\Sigma) + V_2 + V_3 + V_4 \end{aligned}$$

and since $V_4 \subset V_3 \subset V_2$ it follows that $\mathrm{ad}_{\hat{f}}^{L+1}(g) \,\varepsilon\, R^1(\Sigma) \bmod V_2$.

As a second induction assume now that (4.1.3) is true for $k = 1,\ldots,K$. Then

$$\begin{aligned}\mathcal{R}^{K+1}(\hat{\Sigma}) &= [\mathcal{R}^1(\hat{\Sigma}), \mathcal{R}^K(\hat{\Sigma})] \\ &= [\mathcal{R}^1(\Sigma), \mathcal{R}^K(\Sigma)] \bmod ([V_2\ \mathcal{R}^K(\Sigma)] + [\mathcal{R}^1(\Sigma), V_{K+1}] \\ &\qquad + [V_2, V_{K+1}]) \\ &= \mathcal{R}^{K+1}(\Sigma) \bmod V_{K+2}\end{aligned}$$

the last identity following since $V_K \supset \mathcal{R}^K$, thereby establishing (4.1.3) for $k = K+1$.

Since $\mathcal{R}^{p+1}(\Sigma) = V_{p+1} = \{0\}$ by the properties of the g.p.f., it follows that $\mathcal{R}^{p+1}(\hat{\Sigma}) = 0$ and $\mathcal{R}^p(\hat{\Sigma}) = \mathcal{R}^p(\Sigma)$. Thus, we find

$$\mathcal{S}^k(\hat{\Sigma}) = \mathcal{R}^p(\hat{\Sigma}) + \ldots . + \mathcal{R}^k(\hat{\Sigma}) = (\mathcal{R}^p(\Sigma) + \ldots + \mathcal{R}^k(\Sigma)) \bmod V_{k+1}$$

and in particular, $\mathcal{S}^k(\hat{\Sigma}) + V_{k+1} = S^k(\Sigma) + V_{k+1} = \mathcal{R}^k(\Sigma) + V_{k+1}$. Now

$$\mathcal{S}^{k-1}(\Sigma)(0) \big/ \mathcal{S}^k(\Sigma)(0) = (\mathcal{R}^{k-1}(\Sigma)(0) + \mathcal{S}^k(\Sigma)(0)) \big/ (\mathcal{R}^k(\Sigma)(0) + \mathcal{S}^{k+1}(\Sigma)(0))$$

which, since Σ is assumed to be g.c.p.f, yields that

$$\begin{aligned}\mathcal{S}^{k-1}(\Sigma)(0) \big/ \mathcal{S}^k(\Sigma)(0) &= (\mathcal{R}^{k-1}(\Sigma)(0) + V_k(0)) \bmod (\mathcal{R}^k(\Sigma)(0) + V_{k+1}(0)) \\ \text{(4.1.4)} \qquad &= (\mathcal{S}^{k-1}(\hat{\Sigma})(0) + \mathcal{S}^k(\Sigma)(0)) \big/ (\mathcal{S}^k(\hat{\Sigma})(0) + \mathcal{S}^{k+1}(\Sigma)(0)).\end{aligned}$$

We show by induction that $\mathcal{S}^k(\hat{\Sigma})(0) = V_k(0)$. This is certainly true for $k = p$ since then

$$\mathcal{S}^p(\hat{\Sigma})(0) = \mathcal{R}^p(\hat{\Sigma})(0) = \mathcal{R}^p(\Sigma)(0) = V_p(0)$$

so we assume it to be true for $k \leq j \leq p$. But, by (4.1.4) we have

$$\begin{aligned}\mathcal{S}^{k-1}(\Sigma)(0) \big/ \mathcal{S}^k(\Sigma)(0) &= (\mathcal{S}^{k-1}(\hat{\Sigma})(0) + \mathcal{S}^k(\Sigma)(0)) \big/ \mathcal{S}^k(\Sigma)(0) \\ &= \mathcal{S}^{k-1}(\hat{\Sigma})(0) \big/ \mathcal{S}^k(\Sigma)(0)\end{aligned}$$

and by the induction hypothesis $\mathcal{S}^{k-1}(\hat{\Sigma})(0) \supset \mathcal{S}^k(\hat{\Sigma})(0) = \mathcal{S}^k(\Sigma)(0)$. Thus

$$\mathcal{S}^{k-1}(\hat{\Sigma})(0) = \mathcal{S}^{k-1}(\Sigma)(0) = V_{k-1}(0)$$

as required. By definition of g.c.p.f. it follows that $\hat{\Sigma}$ is of the desired form.

The second part of this theorem is proved in similar fashion, using instead the subspaces $\{0^k\}$ of each observation space defined as the linear span of functions of the form

$$L_f^{k_o} L_{g_{i_1}} L_f^{k_1} \ldots\ldots L_{g_{i_k}} L_f^{k_k}(h)$$

for $k_\alpha \geq 0$, $1 \leq i_\beta \leq m$. Then each subspace $\hat{H}^k$ determining the g.o. structure can be written as

$$(4.1.5) \qquad \hat{H}^k = 0^r + 0^{r-1} + \ldots + 0^{k+1} + 0^k.$$

This is easily seen, for if we denote the r.h.s. by $\bar{H}^k$ then $\{\bar{H}^k\}$ is a descending chain of subspaces satisfying the invariance conditions $L_f(\bar{H}^k) \subset \bar{H}^k$, $L_{g_i}(\bar{H}^k) \subset \bar{H}^{k+1}$. By the remarks following Th^m (2.1.1) it follows that $\bar{H}^k \supset \hat{H}^k$. Conversely, we may inductively define 0^k as

$$0^{\bar{o}} \quad = \mathrm{Sp}\{L_f^k(h);\ k \geq 0\}$$

$$(4.1.6) \qquad 0^{k+1} \quad = \{L_f^j L_{g_i}(\phi);\ j \geq 0,\ 1 \leq i \leq m \text{ and } \phi \in 0^k\}$$

from which we see that $\bar{H}^k \subset \hat{H}^k$.

We next claim that, in similar vein to (4.1.3),

$$(4.1.7) \qquad 0^k(\hat{\Sigma}) = 0^k(\Sigma) \bmod Q^{r-k-2}$$

First note that $h \in 0^o(\hat{\Sigma}) \cap 0^o(\Sigma)$. So assume that $\phi \in 0^o(\hat{\Sigma})$ and takes the form $\phi = \phi_1 + \phi_2$ with $\phi_1 \in 0^o(\Sigma)$, $\phi_2 \in Q^{r-2}$. Then

$$L_{\hat{f}}(\phi) = L_f(\phi_1 + \phi_2) - L_{\frac{1}{2}g'}(\phi_1 + \phi_2)$$

But, by the properties of the g.p.f, $L_f : Q^{r-2} \to Q^{r-2}$ and $0^o(\Sigma) \subset Q^r$. Hence

$$L_{\hat{f}}(\phi) = L_f(\phi_1) + \psi_1$$

with $\psi_1 \in Q^{r-2}$ as required, and so (4.1.7) is established for the case $k = 0$. Now suppose it to be true for $k = 0,\ldots,J$. Then by the inductive definition (4.1.6) every function in $0^{J+1}(\hat{\Sigma})$ can be written as $L_f^j(L_{g_i}(\phi_1 + \phi_2))$ with $\phi_1 \in 0^J(\Sigma)$ and $\phi_2 \in Q^{r-J-2}$. From linearity of the

Lie derivative and since $L_{g_i} \varepsilon \ V_1$ it follows that

$$L^{j}_{\hat{f}} \ (L_{g_i}(\phi_1 + \phi_2)) \ \varepsilon \ L^{j}_{\hat{f}}(L_{g_i}(\phi_1)) + Q^{r-(J+1)-2}$$

and a simple induction shows that,

$$L^{j}_{\hat{f}}(L_{g_i}(\phi_1)) \ \varepsilon \ L^{j}_{f}(L_{g_i}(\phi_1)) + Q^{r-(J+1)-2}$$

as required.

[Before proceeding further, we make the remark that (4.1.7) shows that the Volterra series describing $\hat{\Sigma}$ has the same length as that of Σ since $0^r(\hat{\Sigma}) = 0^r(\Sigma) = \mathbb{R}$].

We now show that $\hat{\Sigma}$ is in g.o.p.f. By (4.1.5) and (4.1.7) we see that

$$\begin{aligned} d\hat{H}^k(\hat{\Sigma}) &= d0^r(\hat{\Sigma}) + .. + d0^k(\hat{\Sigma}) \\ &= [d0^r(\Sigma) + .. + d0^k(\Sigma)] \bmod d\ Q^{r-k-2} \end{aligned}$$

and, in particular,

$$\begin{aligned} d\hat{H}^k(\hat{\Sigma}) + dQ^{r-k-2} &= d\hat{H}^k(\Sigma) + dQ^{r-k-2} \\ &= d0^k(\Sigma) + d0^{k+1}(\Sigma) + dQ^{r-k-2} \end{aligned}$$

But then

$$d\hat{H}^{k-1}(\Sigma)(0) \Big/ d\hat{H}^k(\Sigma)(0) = (d0^{k-1}(\Sigma)(0) + d0^k(\Sigma)(0) + d\hat{H}^{k+1}(0)) \Big/ (d0^k(\Sigma)(0) + d0^{k+1}(\Sigma)(0) + d\hat{H}^{k+2}(0))$$

which, since Σ is in g.o.p.f (so that $d\hat{H}^k(\Sigma)(0) = dQ^{r-k}(0)$) gives

$$\begin{aligned} d\hat{H}^{k-1}(\Sigma)(0) \Big/ d\hat{H}^k(\Sigma)(0) &= d0^{k-1}(\Sigma)(0) + d0^k(\Sigma)(0) + dQ^{r-k}(0)) \Big/ d0^k(\Sigma)(0) + d0^{k-1}(\Sigma)(0) + dQ^{r-k}(\Sigma)(0) \\ &= (d\hat{H}^{k-1}(\hat{\Sigma})(0) + dQ^{r-k}(0)) \Big/ (d\hat{H}^k(\hat{\Sigma})(0) + dQ^{r-k}(0)) \end{aligned}$$

Inductively, we then find that $d\hat{H}^k(\hat{\Sigma})(0) = dQ^{r-k}(0)$. Indeed, for $k = r$ this result is true from (4.1.7) and the definition of $\hat{H}^k$. Similarly, for $k = r-1$, we have

$$d\hat{H}^{r-1}(\hat{\Sigma})(0) = dO^{r-1}(\hat{\Sigma})(0) + dO^{r}(\hat{\Sigma})(0) = dO^{r-1}(\Sigma)(0) + dO^{r}(\Sigma)(0)$$
$$= d\hat{H}^{r-1}(\Sigma)(0).$$

Let us assume, therefore, that it is true for $j \geq k$. Then

$$d\hat{H}^{k-1}(\Sigma)(0) \Big/ d\hat{H}^{k}(\Sigma)(0) = \left(d\hat{H}^{k-1}(\hat{\Sigma})(0) + d\hat{H}^{k+1}(\Sigma)(0)\right) \Big/ d\hat{H}^{k}(\hat{\Sigma})(0)$$

But, by induction, $d\hat{H}^{k-1}(\Sigma)(0) \supset d\hat{H}^{k}(\hat{\Sigma})(0) \supset d\hat{H}^{k}(\Sigma)(0) \supset d\hat{H}^{k+1}(\Sigma)(0)$

and hence,

$$d\hat{H}^{k-1}(\hat{\Sigma})(0) = d\hat{H}^{k-1}(\Sigma)(0)$$

Thus showing (as required) that $\hat{\Sigma}$ is in g.o.p.f.

□

Remark: It should be noted that the proof of this theorem actually shows that if X is any vector field in V_2 then

$\{f,g,h\}$ in g.c.p.f. (resp. g.o.p.f) => $\{f + X,g,h\}$ is in g.c.f.(resp. g.o.p.f) since the fundamental identities (4.1.3) and (4.1.7) follow by exactly the same arguments.

□ .

As we pointed out earlier, the prime significance of this theorem is the natural corollory that if $\hat{\Sigma}$ is minimal and in g.p.f. or Σ is minimal and in both g.o.p.f. and g.c.p.f., then the tensor algebra $\Omega^{o}(\hat{\Sigma}) \equiv W_n$, where n is the dimension of the state space of Σ and $\hat{\Sigma}$. Consequently if it can be shown that the Estimation Algebra of such a system is equal to $\Omega^{o}(\hat{\Sigma})$, then from Th^{m} (4.0.1) it follows that Σ will have no non-trivial f.d.c. statistics. That this is possible was shown by example in the introduction, but before we go on to analyse the situation further, we remark that $\Omega^{o}(\hat{\Sigma})$ will always be homomorphic to W_m, for some m, even if $\hat{\Sigma}$ is non-minimal, provided that $\hat{\Sigma}$ is in g.p.f. (for instance if Σ is itself only in g.p.f.). Indeed, the maps β,γ defined in Theorem (1.2.6) extend in a natural, homomorphic fashion to maps $\hat{\beta}:\mathscr{X}_A(\hat{\Sigma})\to\mathscr{X}_A(\tilde{\Sigma})$ and $\gamma:\mathscr{U}(\mathscr{S}(\hat{\Sigma}))\to\mathscr{U}(\mathscr{S}(\tilde{\Sigma}))$ where $\tilde{\Sigma}$ is a minimal, g.p.f. of $\hat{\Sigma}$. If we let

$\pi = \hat{\beta} \otimes \hat{\gamma}$ then π is clearly a linear map from $\Omega^{o}(\hat{\Sigma})$ onto $\Omega^{o}(\tilde{\Sigma})$. Moreover, if $\phi X, \psi Y$ are elements of $\mathscr{H}(\hat{\Sigma}) \otimes \mathscr{S}(\hat{\Sigma})$, then

$$\begin{aligned}\pi([\phi X,\psi Y]) &= \pi\,(\phi\psi[X,Y] + \phi L_X(\psi)Y - \psi L_Y(\phi)X)\\ &\triangleq \beta(\phi)\beta(\psi)[\gamma(X),\gamma(Y)] + \beta(\phi)L_{\gamma(X)}\beta(\psi)\psi(Y)\\ &\qquad\qquad - \beta(\psi)L_{\psi(Y)}\beta(\phi)\gamma(X)\\ &= [\pi(\phi X),\pi(\psi Y)]\end{aligned}$$

(here we have made use of the identity a) given in Lemma (3.1.2) to expand $[\phi X,\psi Y]$). Inductively, it follows that π is also a homomorphism.

From the properties of the g.p.f. this leads us to the following deduction:

(4.1.8) If $\Lambda(\Sigma) = \Omega^{o}(\hat{\Sigma})$ and $\hat{\Sigma}$ is in g.p.f. on $\mathbb{R}^n$ then $\Lambda(\Sigma)$ is a Lie subalgebra of W_n and is epimorphic to W_m for some $m \leq n$.

This clearly has implications for the algebraic estimation properties of Σ - for instance, if ϕ: $\Lambda \to \Lambda_1$ is a Brockett homomorphism so $\Lambda_1 \subset \Gamma^{\omega}(TM)$ with $\ker \pi \subset \ker \phi$, then ϕ must be trivial (otherwise $\bar{\pi}$: $W_m \to \Lambda_1$ defined by $\bar{\pi}(X) = \phi(\pi^{-1}(X))$ is a non trivial homomorphism contradicting $\mathrm{Th}^{\underline{m}}$ (4.0.1)). However, the full extent of this influence has yet to be determined.

§4.2. Drift Independent Observability

In this section we make further preparations for our generalisation of the Hazewinkel-Marcus example by developing a canonical representation for systems satisfying a strong observability condition. This form was inspired by a description given, first by Gauthier and Bornard [1] and subsequently (in more elegant terms) by Nijmeier [1], in response to the observation that linear systems are observable for any input. In particular, this means that the initial condition, and consequently the state, can be reconstructed through knowledge of the output derivatives $\{y(0), y'(0), \ldots, y^{(n-1)}(0)\}$, independently of control. By assuming, therefore, the linear analytic system $\Sigma = \{f, g_i, h_j ; 1 \leq j \leq m\}$, defined on a manifold M^n with output in $\mathbb{R}$, to be observable for any constant input the following local description (valid for any smooth input) was derived by the above authors

$$(4.2.1) \qquad \begin{cases} \dot{z} = \begin{bmatrix} \dot{z}_1 \\ \cdot \\ \cdot \\ \dot{z}_j \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \dot{z}_n \end{bmatrix} = \begin{bmatrix} z_2 \\ \cdot \\ \cdot \\ z_{j-1} \\ \cdot \\ \cdot \\ z_{n-1} \\ F(z) \end{bmatrix} + \sum_{i=1}^{m} u_i(t) \begin{bmatrix} g_{1j}(z_1) \\ \cdot \\ \cdot \\ g_{ji}(z_1, \ldots, z_j) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ g_n(z_1, \ldots, z_n) \end{bmatrix} \\ y = z_n \end{cases}$$

(For multiple outputs the description is more complicated, relying on a decomposition of the state vector according to a set of "dual observability indices", but retaining a structure similar to that of the companion forms in linear theory. We refer to Nijmeier [1] for details).

The above idea of input-independent observability has a certain intuitive appeal for the filtering problem since we could argue that we can assume the (smooth) input is an approximation to the random driving force and still be able to determine the state by using the Doss-Sussmann concept of a solution to the stochastic system. Further evidence to corroborate this argument is given by the Hazewinkel-Marcus example, for which the underlying system is

$$(4.2.2) \qquad \begin{cases} \begin{cases} \dot{x}_1 = u \\ \dot{x}_2 = x_1 + ux_1 \end{cases} \\ y = x_2 \end{cases}$$

and is thus observable. Indeed, if we define $f = x_1 \frac{\partial}{\partial x_1}$, $g = \frac{\partial}{\partial x_i} + x_1 \frac{\partial}{\partial x_2}$, $h = x_2$ then, at all points $x \varepsilon \mathbb{R}^2$,

$$T_x^* \mathbb{R}^2 = d\mathscr{H}(x) = Sp\{dh(x), dL_f(h)(x)\}$$

$$= Sp\{dh(x), dL_g(h)(x)\}.$$

Moreover, f is clearly in the companion form of system (4.2.1), yet (4.2.2) is not input-independent observable. For, if $u = -1$ then $\dot{x}_2 = 0$ and so we obtain an unobservable system (we remark, though, that the system is observable for all $u \neq -1$ thus agreeing with the theorem stating that observable linear analytic systems are observable for almost all smooth inputs, Sussmann [5]).

We see, therefore, that despite Sussmann's result, observability can depend on input, and it is this point of view which we wish to develop in this section. We begin with a definition

DEFINITION 4.2.1

Let $\Sigma = \{f,g,h\}$ be an accessible, weakly observable linear analytic system defined on a manifold M^n with output in $\mathbb{R}$. Then Σ is said to be drift independent observable (d.i.o) at $x_1 \in M$ if the system

$$(4.2.3) \qquad \left| \begin{array}{l} \dot{z} = g(z) \\ \\ \tilde{y} = h(z) \end{array} \right.$$

is weakly observable at x_1. In particular, this means that the codistribution generated by $\{L_g^k(h);\ 0 \leq k \leq n-1\}$ spans T_x^*M, for all x in some neighbourhood of x_1 in M. The system is d.i.o. if it is di.o at x_1 for all x_1 in M.

□

Let Σ now be a d.i.o. system as in Def^n(4.2.1) and consider the map $\Phi: M \to \mathbb{R}^n$ with i^{th}-component

$$\Phi_i(x) = L_g^{n-i}(h)(x) \qquad i = 1, \ldots, n.$$

Then, by the definition of drift-independent observability and the inverse function theorem it follows that Φ is a diffeomorphism of a neighbourhood U of some point x_1 onto a neighbourhood of $\Phi(x_1)$ in $\mathbb{R}^n$. We thus obtain a local description of the system Σ by setting $z(t) = \Phi(x(t))$, where $x(t)$ is a trajectory of Σ. This transformation results in the representation on $\Phi(U)$

$$\dot{z}(t) = \Phi_*\dot{x}(t) = (\Phi_* f)(z(t)) + u(t)(\Phi_* g)(z(t))$$

But from the results given in §1.1 we know that

$$\begin{aligned}(\Phi_* g)_i(z) &= L_g(\Phi_i)(\Phi^{-1}(z)) \\ &= L_g^{\,n-i+1}(h)(\Phi^{-1}(z)) \\ &= \begin{cases} z_{i-1} & 2 \leq i \leq n \\ \tilde{g}_1(z) & i = 1 \end{cases}\end{aligned}$$

where $\tilde{g}_1(z) = L_g^n(h)(\Phi^{-1}(z))$. Thus, on $\Phi(U)$, by defining $\tilde{f} = \Phi_* f$ we see the system can be represented

$$\begin{cases} \dot{z} = \tilde{f}(z) + u(t)\begin{bmatrix} \tilde{g}_1(z) \\ z_1 \\ \vdots \\ z_{n-1} \end{bmatrix} = \tilde{f}(z) + u(t)\tilde{g}(z) \\ y(t) = z_n(t) \end{cases}$$

Of course the prime example of a system in d.i.o. form is that of (4.2.2), but we include here a simple example of the construction outlined above to motivate our next steps. So consider the system on $\mathbb{R}$;

$\{f = ax\frac{\partial}{\partial x},\ g = b\frac{\partial}{\partial x},\ h(x) = \frac{x^3}{3} + x\}$ or

$$(4.2.4) \qquad \begin{cases} \dot{x} = ax + bu & b \neq 0 \\ y = \dfrac{x^3}{3} + x \end{cases}$$

Then this system is d.i.o. since $h(x)$ is a diffeomorphism, we therefore set $z(t) = h(x(t))$ to obtain first

$$(4.2.5) \qquad \dot{z} = ax(x^2+1) + ub(x^2+1).$$

By appealing to Cardans technique for obtaining the roots of cubic equations, we next find

$$x(t) = h^{-1}(z)(t)) = \left[\frac{3z+\sqrt{(9z^2+4)}}{2}\right]^{1/3} + \left[\frac{3z-\sqrt{(9z^2+4)}}{2}\right]^{1/3}$$

which, on substitution into (4.2.5) yields the canonical form for (4.2.4)

$$(4.2.6) \qquad \dot{z} = a\, h^{-1}(z)(h^{-1}(z))^2 + 1) + u\, b(h^{-1}(z)^2 + 1)$$

$$y(t) = z(t)$$

Clearly, then, under these most general hypotheses there is little more which can be said on the structure of the transformed vector fields, $\tilde{f}(z)$ and $\tilde{g}(z)$ even if the original system has a fairly simple description (note that (4.2.4) is actually in g.p.f. and is minimal but (4.2.6) no longer even has polynomial dynamics). For this reason we make some specialising assumptions on the system Σ, namely that it is a minimal g.p.f. with state space $\mathbb{R}^n = \bigoplus_{i=1}^{n} \mathbb{R}^{n_i}$, each $n_i = 1$, and $h\varepsilon Q^n$. There are two immediate consequences of these conditions; first, since $\tilde{g}_1(z) = L^n_{\tilde{g}}(h)$ it follows that $\tilde{g}_1(z)$ is constant, (which in the sequel is shown to be non-zero and, hence, can be normalised to 1) and, secondly, the mapping Φ is polynomial so, by d.i.o. and Palais' GIFT, it follows that it is in fact a global diffeomorphism of $\mathbb{R}^n$. Hence, the d.i.o. form of such a system is given by

$$\dot{z} = \tilde{f}(z) + u(t)\begin{bmatrix} 1 \\ z_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ z_{n-1} \end{bmatrix} \qquad z\varepsilon\mathbb{R}^n$$

Following on from this development, it is natural to ask whether the graded structure is preserved. As we have seen in §1.3, this hinges on whether or not Φ has a polynomial inverse. The following results shows that under the above hypotheses this is always the case.

THEOREM 4.2.2

Let $\Sigma = \{f,g,h\}$ be a minimal system in g.p.f. with respect to the decomposition $\mathbb{R}^n = \bigoplus_{i=1}^{n} \mathbb{R}$ and d.i.o. at $x_o\varepsilon\mathbb{R}^n$. Then the system is d.i.o. and the corresponding canconical description is also in g.p.f. with respect to the same graded structure.

Proof

We first show that d.i.o. at a point $\Rightarrow$ d.i.o. everywhere. By assumption, $g\varepsilon V_1$ and the specific graded structure further implies that Φ takes the form

$$\Phi(x) = \begin{bmatrix} \Phi_1(x_1) \\ \cdot \\ \cdot \\ \Phi_i(x_1,\ldots,x_i) \\ \cdot \\ \cdot \\ \cdot \\ \Phi_n(x_1,\ldots,x_n) \end{bmatrix} \tag{4.2.7}$$

Moreover, each $\Phi_i \varepsilon Q^i$ and is thus at most linear in x_i, and independent of $x_{i+1},\ldots,x_n$. So, with respect to these coordinates, we find

$$D\Phi_x = \begin{bmatrix} \partial\Phi_1/\partial x_1 & & 0 & \cdot\ \cdot\ \cdot & 0 \\ & \cdot & & & \vdots \\ & & \cdot & & \vdots \\ \tilde{\Phi}(x_1,\ldots,x_{n-1}) & & & \cdot & 0 \\ & & & \cdot & \dfrac{\partial\Phi_n}{\partial x_n} \end{bmatrix}$$

In particular, it follows that for all $x\varepsilon\mathbb{R}^n$

$$\det D\Phi_x = \prod_{i=1}^{n} \frac{\partial\Phi_i}{\partial x_i} = \text{a constant, } c$$

and c is non-zero by virtue of d.i.o. at x_o. $D\Phi_x$ is therefore a unimodular matrix so we conclude immediately that Σ is d.i.o. everywhere and Φ is a diffeomorphism of $\mathbb{R}^n$.

We now turn our attention to showing that the d.i.o. form of Σ is also in g.p.f. To do this we only need show that $\tilde{f}\varepsilon V_o$, ie. that

$$\tilde{f}(z) = \sum_{i=1}^{n} \tilde{f}_i(z)\frac{\partial}{\partial z_i}\text{, with } \tilde{f}_i \ \varepsilon\ Q^i[z] \tag{4.2.8}$$

where $Q^i[z]$ now denotes the space of polynomials in z of weight $\leq i$ with respect to the gradation $\bigoplus_{i=1}^{n} \mathbb{R}$, since the input vector field takes the form

$$\tilde{g} = \begin{bmatrix} \tilde{g}_1 \\ z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_n \end{bmatrix}$$

which is an element of V_1 w.r.t. this gradation. But from (4.2.7) we see that

$$z_1 = \Phi_1(x_1)$$

and $\Phi_1 \in Q^1[x]$ therefore takes the form $\Phi_1(x_1) = \alpha_1 x_1 + \beta_1$. But, as Φ is a diffeomorphism, it follows that $\alpha_i \neq 0$ and hence

$$x_1 = \frac{z_1 - \beta_1}{\alpha_1}$$

ie. $x_1 = \Psi_1(z)$ with $\Psi_1 \in Q^1[z]$. We inductively assume that for $1 \leq i \leq k$ each coordinate x_i can be expressed as

$$x_i = \Psi_i(z_1,\ldots,z_i) \quad , \quad \Psi_i \in Q^i[z].$$

Then, as above, since $\Phi_{k+1} \in Q^{k+1}[z]$ and $\frac{\partial\Phi}{\partial x_{k+1}} \neq 0$ it follows that

$$z_{k+1} = \alpha_{k+1}\, x_{k+1} + \tilde{\Phi}_{k+1}(x_1,\ldots,x_k), \ \tilde{\Phi}_{k+1} \in Q^{k+1}[x].$$

or

$$x_{k+1} = \frac{1}{\alpha_{k+1}}\, z_{k+1} + \tilde{\Phi}_{k+1}\,(\Psi_1(z_1),\ldots,\Psi_k(z_1,\ldots,z_k))$$

$$= \frac{1}{\alpha_{k+1}}\, z_{k+1} + \tilde{\Psi}_{k+1}(z_1,\ldots,z_k) = \Psi_{k+1}(z)$$

Clearly, $\tilde{\Psi}_{k+1}$ is a polynomial so to conclude the induction it remains to prove that it is of weight $\leq k+1$. This is perhaps easiest seen if we first consider $\tilde{\Psi}_{k+1}$ as the polynomial

$$\tilde{\Psi}_{k+1}(z) = \tilde{\Phi}_{k+1}(\Psi_1,\ldots,\Psi_k) = \sum_{w_k(\alpha)=0}^{k+1} A_\alpha\, \Psi_1^{\alpha_1} \ldots \Psi_k^{\alpha_k}$$

$$\text{s.t. } w_k(\alpha) \triangleq \sum_{m=1}^{k} m\alpha_m \ .$$

But $\Psi_i \in Q^i[z]$, by assumption, so $\Psi_i^{\alpha_i} \in Q^{i\alpha_i}[z]$ and hence each product $\Psi_1^{\alpha_1} \ldots \Psi_k^{\alpha_k} \in Q^{w_k(\alpha)}[z]$. In particular, it follows that $\tilde{\Psi}_{k+1} \in Q^{k+1}[z]$, completing the induction.

We have therefore shown that not only does Φ have a polynomial inverse but also that this inverse takes the form

$$\Phi^{-1}(z) = \begin{bmatrix} \Psi_1(z_1) \\ \vdots \\ \Psi_n(z_1,\ldots,z_n) \end{bmatrix} \text{with } \Psi_i \in Q^i[z]$$

By definition of the vector field $\tilde{f}$ it now follows that each component, $\tilde{f}_i$, as defined in (4.2.8) and also satisfying $\tilde{f}_i(z) \triangleq L_f(\Phi_i)(\Phi^{-1}(z))$, is polynomial. Moreover, since $f \in V_o$ we must have $L_f(\Phi_i) \in Q^i[x]$, so using the same argument as before we deduce that $\tilde{f}_i \in Q^i[z]$ thus showing that $\tilde{f} \in V_o$ as required. Further, since the d.i.o. is now seen to be in g.p.f. it follows from Th^{m} (1.3.1) (iv) that $\tilde{g}_1 \neq 0$.

□

As indicated previously, the concept of drift independent observability of g.p. forms is fundamental to our generalisation of the Hazelwinkel-Marcus example thus, as in §4.1, we need to know how such systems behave under perturbation by the Itô correction term. Clearly, the d.i.o. rank condition will remain unaffected since the control vector field g and the output function h are unchanged. To prove that minimality is also preserved we show that the system is actually in g.c.p.f. and then apply Th^{m} (4.1.1), but to do so we first need the following result which plays a further, important role in the next section.

<u>THEOREM 4.2.3</u>

Let X be the vector field on $\mathbb{R}^n \left\{\frac{\partial}{\partial x_1} + x_1 \frac{\partial}{\partial x_2} + x_2 \frac{\partial}{\partial x_3} + \ldots + x_n + \frac{\partial}{\partial x_n}\right\}$.

Then

(i) L_X is a surjective automorphism of $\mathbb{R}[x_1,\ldots,x_n]$

(ii) With respect to the gradation $\oplus^n\mathbb{R}$ of $\mathbb{R}^n$, $L_X:Q^k\to Q^{k-1}$ and $ad_X:V_k\to V_{k+1}$ are surjections.

Proof

We need only show that $L_X:Q^k\to Q^{k-1}$ is a surjection, the other claims being immediate corollaries since

a) $\mathbb{R}[x_1,\ldots,x_n] = \bigoplus_{k\geq 0} Q^k$

and b) $V_k = \bigoplus_{j=k}^{n} Q^{j-k} \oplus \Delta_j$

so $L_X(\mathbb{R}[x_1,\ldots,x_n]) = \bigoplus_{k\geq 0} L_X(Q^k) = \bigoplus_{k\geq 1} Q^{k-1} = \mathbb{R}[x_1,\ldots,x_n]$

proving (i). Similarly, if $Y\varepsilon V_k$ it can be written as $Y = \sum_{j=1}^{n}\phi_j(x)\frac{\partial}{\partial x_j}$ with each coordinate $\phi_j \ \varepsilon\ Q^{j-k}$. Then, defining $x_o = 1$, we have

$$\begin{aligned} ad_X(Y) &= [\sum_{i=1}^{n} x_{i-1}\frac{\partial}{\partial x_i},\ \sum_{j=1}^{n}\phi_j(x)\frac{\partial}{\partial x_j}] \\ &= \sum_{j=1}^{n} L_X(\phi_j)\frac{\partial}{\partial x_j} - \Sigma\phi_j[\frac{\partial}{\partial x_j}, x_{i-1}]\frac{\partial}{\partial x_i} \\ &= \sum_{j=1}^{n}(L_X(\phi_j) - \phi_{j-1})\frac{\partial}{\partial x_j} \quad \varepsilon \quad V_{k+1} \end{aligned}$$

where $\phi_o = 0$. To show surjectivety of ad_X, we must therefore be able to solve the equations

$$L_X(\phi_1) = \psi_1$$

$$L_X(\phi_j) = \psi_j + \phi_{j-1}$$

for a given set of components $\psi_j \ \varepsilon\ Q^{j-(k+1)}$, $1 \leq j \leq n$. However, since $\phi_j \ \varepsilon\ Q^{j-k}$ this follows trivially from the surjectivity of L_X onto $Q^{j-(k+1)}$.

We prove the surjectivity of L_X as a map from Q^k into Q^{k-1} by showing that $L_X:H^k\to H^{k-1}$ is surjective using induction. For $k = 1$, this is trivial since $\phi\varepsilon H^1 \Rightarrow \phi = \alpha x_1$ for some $\alpha\ \varepsilon\mathbb{R}$, so

$$L_X(\phi) = \sum_{i=1}^{n} x_i \frac{\partial}{\partial x_i} (\phi) = \alpha.$$

But, $H^0 = \mathbb{R}$ and $H^1 \supset Sp\{x_1\}$ so L_X is certainly surjective on H^1. Assume, therefore, that the claim is valid for $k = 1,\dots\dots,K-1$ and suppose that $\{\phi_j\}$ is a basis for H^K. Then we have,

$$\phi_j = x_1^{j_1} \dots\dots x_n^{j_n} \qquad \text{with } w(j) = K.$$

Thus, we can find a basis element, ψ_j, of $H^{K-\ell}$ for which

$$\phi_j = x_\ell \psi_j$$

where ℓ is some integer $1 \leq \ell \leq n$. There are two cases to consider namely (i) $K > n$ and (ii) $K \leq n$.

If $K > n$ then $x_\ell \in H^\ell$ so $\phi_j \in H^\ell \otimes H^{K-\ell}$ with $\ell \leq K-1$. On the other hand, if $K \leq n$, the definition of homogeneity implies that the integers $j_n, \dots, j_{K+1} = 0$ and $j_K \in \{0,1\}$. If $j_K = 0$, then the coordinate x_ℓ chosen is again in H^ℓ with $\ell \leq K-1$ as before giving $\phi_j \in H^\ell \otimes H^{K-\ell}$. But if $j_K = 1$ then by definition $\psi_j = 1$, ie. $\phi_j = x_K$. Thus we find that

$$H^K = \sum_{\ell=1}^{K-1} H^\ell \otimes H^{K-\ell} + Sp\{x_K\}$$

from which the induction follows trivially since

$$\begin{aligned} L_X(H^K) &= \sum_{\ell=1}^{K-1} L_X(H^\ell) \otimes H^{K-\ell} + H^\ell \otimes L_X(H^{K-\ell}) + SpL_X(x_K) \\ &= \sum_{\ell=1}^{K-1} H^{\ell-1} \otimes H^{K-\ell} + \sum_{\ell=1}^{K-1} H^\ell \otimes H^{K-\ell-1} + Sp\,\{x_{K-1}\} \\ &= H^{K-1} + H^{K-1} + Sp\{x_{K-1}\} \\ &= H^{K-1} \end{aligned}$$

as required. The proof for Q^k is then obvious using the decomposition $Q^k = \bigoplus_{\ell=0}^{k} H^\ell$.

□

COROLLARY 4.2.4

Let Σ be the d.i.o. system described in Th^m (4.2.2). Then Σ is in g.c.p.f. and g.o.p.f.

Proof

First note that since $\Phi_*: \mathscr{S}(\Sigma) \to \mathscr{S}(\Phi_*\Sigma)$ is an isomorphism, where Φ is the map defined in (4.2.7), and the graded structures of both Σ and the transformed system $\Phi_*\Sigma$ are identical it follows that we need only show that the d.i.o. form is in g.c.p.f. and g.o.p.f. But now the input vector field $\tilde{g}$ is of the form

$$\tilde{g} = \tilde{g}_1 \frac{\partial}{\partial x_1} + \sum_{i=2}^{n} x_{i-1} \frac{\partial}{\partial x_i}$$

and $\tilde{g}_1 \neq 0$. In particular, the above theorem shows that $\tilde{g}$ acts surjectively on V and therefore induces a surjection A_x: $V_j(x) \to V_{j-1}(x)$ defined by

$$A_x(Y(x)) = \mathrm{ad}_{\tilde{g}}(Y)(x)$$

Now $\Phi_*\Sigma$ is minimal, so in particular $\mathscr{S}(\Phi_*\Sigma)(x) = \mathscr{S}^1(\Phi_*\Sigma)(x) = V_1(x)$. But then

$$V_2(x) \supset \mathscr{S}^2(\Phi_*\Sigma)(x) \supset A_x(\mathscr{S}^1(\Phi_*\Sigma)(x) = A_x(V_1(x)) = V_2(x)$$

and so $\mathscr{S}^2(\Phi_*\Sigma)(x) = V_2(x)$. Inductively it follows that $\mathscr{S}^j(\Phi_*\Sigma)(x) = V_j(x)$ $\forall x \in \mathbb{R}^n$ i.e. $\Phi_*\Sigma$ is in g.c.p.f.

Similarly, $\tilde{g}$ induces a surjection $A^x: W_\ell(x) \to W_{\ell-1}(x)$, where $W_\ell = \bigoplus_{j=1}^{n} Q^{\ell-j} \otimes \Delta^j$ (c.f. Th^m (1.1.2)), defined by

$$A^x(\Sigma\phi_i(x)dx^i) = \Sigma L_{\tilde{g}}(\phi_i)(x)dx^i$$

and since $d\mathscr{H}(x) = d\hat{H}^0(x) = W_n(x)$, the same argument shows that $\Phi_*\Sigma$ is in g.o.p.f. as required.

□

From the above corollary we deduce immediately that the "Itô-perturbed" version of Σ is also minimal.

§4.3. The Estimation Algebra for a Class of D.I.O. Systems

We now come to the main purpose of this chapter, namely the construction of a class of systems for which the estimation algebra is isomorphic to W_n and which contains that studied by Hazewinkel and Marcus. As we have seen, this particular example exhibits several interesting features, for instance it is minimal, d.i.o. and in g.p.f. (indeed is also in d.i.o. form). Unfortunately, the class of all systems with minimal, d.i.o.g.p. realisations must include the scalar linear system

$$\begin{cases} \dot{x} = ax + bu \\ y = x \end{cases}$$

for which the estimation algebra is finite dimensional (c.f. §3.2 example V). So that we must restrict our attention even further. This is achieved by assuming that we can show that the estimation algebra for the d.i.o. form actually contains the operators $L_{\tilde{g}}$, $L_{\tilde{g}}^2$ and z_n^2, where throughout this section $\tilde{g}$ is the vector field $\sum_{i=1}^{n} z_{i-1} \frac{\partial}{\partial z_1}$, $z_o = 1$, and that $n \geq 2$. Thus, it is our intention to prove

THEOREM 4.3.1

Suppose that the underlying system of

$$\begin{cases} dz_1 = f_1(z_1)dt + dw \\ dz_i = f_i(z_1,..,z_i)dt + z_{i-1}dw \qquad 2 \leq i \leq n \end{cases}$$

$$dy = z_n dt + dv$$

is defined on $\mathbb{R}^n$ and is in minimal g.p.f. w.r.t. the gradation $\oplus^n \mathbb{R}$. Assume further that $n \geq 2$ and the estimation algebra contains $\{L_{\tilde{g}}, L_{\tilde{g}}^2, z_n^2\}$. Then $\Lambda = W_n$

$\square$

We prove this result in two stages, by first showing that $\Lambda(\Sigma)$ contains $\mathbb{R}[z_1,\ldots,z_n]$ and then using the surjectivity of $L_{\tilde{g}}$ combined with the algebraic structure properties of g.p.f's. Before doing so, however, we remark that if Σ_x and Σ_z are the two Itô systems on $\mathbb{R}^n$

$$\Sigma_x = \begin{cases} dx = f(x)dt + g(x)dw_t \\ dy = h(x)dt + dv_t \end{cases} \qquad \Sigma_z = \begin{cases} dz = \bar{f}(z)dt + \bar{g}(z)dw_t \\ dy = \bar{h}\,(z)dt + dv_t \end{cases}$$

and there is a diffeomorphism $\alpha:\mathbb{R}^n \to \mathbb{R}^n$ such that $\alpha(x_t) = z_t$, then α induces an isomorphism from $\Lambda(\Sigma_x)$ to $\Lambda(\Sigma_z)$, Brockett [3]. Consequently, if Σ_x is a minimal d.i.o. g.p.f. with respect to $\oplus^n \mathbb{R}$ and $\{L_g, L_g^2, h^2\} \subset \Lambda(\Sigma_x)$, then the coordinate transformation Φ used in Th^m (4.2.2) to construct the d.i.o. form induces an isomorphism between $\Lambda(\Sigma_x)$ and W_n provided we can show that Σ_z is in minimal g.p.f. This is not quite as obvious as it sounds since we have to use Itô calculus to derive the dynamics. Thus, whilst $\bar{g} = \Phi_* g = \tilde{g}$ and $\bar{h} = h \circ \Phi^{-1} = z_n$, the drift vector field $\bar{f}$ is given by the components

$$\bar{f}_i(z) = \Sigma f_j \left.\frac{\partial \Phi_i}{\partial_j}\right|_{\Phi^1(z)} + \tfrac{1}{2}\,\Sigma\, g_k g_j \left.\frac{\partial^2 \Phi_i}{\partial x_k \partial x_j}\right|_{\Phi^{-1}(z)}.$$

$$= L_f(\Phi_i)(\Phi^{-1}(z)) + \tfrac{1}{2}\, L_g^2(\Phi_i)(\Phi^{-1}(z))$$

or

$$\bar{f}(z) = \Phi_* \hat{f} + \sum_{i=2}^{n} z_{i-2} \frac{\partial}{\partial z_i}$$

where, as before, $\hat{f} = f - \frac{1}{2}\frac{dg}{dx}g$. Now, from Cor^y (4.2.4) we know that $\{f,g,h\}$ is in symmetric p.f. Consequently, using Th^m (4.1.1) we find that $\{\hat{f},g,h\}$, and hence $\{\Phi_*\hat{f},\bar{g},\bar{h}\}$, is also in symmetric p.f. In particular, we can apply the remarks following Th^m (4.1.1) to see that $\{\bar{f},\bar{g},\bar{h}\}$ is in minimal g.p.f. since $\sum_{i=2}^{n} z_{i-2} \frac{\partial}{\partial z_i} \in V_3 \subset V_2$. We have therefore shown that Σ_z is in the form required for Th^m (4.3.1) to apply. Moreover, by assumption on $\Lambda(\Sigma_x)$, we know that $\{L_{\Phi_* g}, L^2_{\Phi_* g}, (h\circ\Phi^{-1})^2\} = \{L_{\tilde{g}}, L^2_{\tilde{g}}, z_n\} \subset \Lambda(\Sigma_z)$

and so $\Lambda(\Sigma_x) \approxeq W_n$ as required.

We now turn our attention to the proof of Th^m (4.3.1) which, as we said, is in two parts.

LEMMA 4.3.2

Under the conditions of Th^m (4.3.1), $\mathbb{R}[z_1,\ldots,z_n] \subset \Lambda$

Proof

Since $\Lambda \triangleq \{F,G\}_{L.A.}$ and $G = z_n$, the assumption that $L_{\underset{\sim}{g}} \in \Lambda$ and the identity $[L_{\underset{\sim}{g}}, z_i] = z_{i=1}$ trivially imply that $\{1,z_1,\ldots,z_n\} \subset \Lambda$. Similarly the inductive application of the equation

$$[[L_g^2, z_k]z_j] = z_{k-1}z_{j-1}$$

yields the cross products $\{z_n^2, z_k z_j;\ 0 \leqslant k \leqslant n-1,\ 0 \leqslant j \leqslant n-1\} \subset \Lambda$. It now follows that

$$\{\mathbb{R}[z_1,\ldots,z_{n-1}], z_n L_g\} \subset \Lambda. \tag{4.3.1}$$

To see that this is true, we show first that $L_{\underset{\sim}{g}}^m \in \Lambda\ \forall m \geqslant 1$, which hypothesis is known to hold for $m = 1,2$, and so is assumed to hold for $m = 1,\ldots,M-1$. Then

$$z_1 L_{\underset{\sim}{g}}^{m-1} = \frac{1}{2m}[L_{\underset{\sim}{g}}^m, z_1] - (m-1)L_{\underset{\sim}{g}}^{m-2} \in \Lambda \text{ for } 1 \leqslant m \leqslant M-1 \tag{4.3.2}$$

and

$$[L_{\underset{\sim}{g}}^{M-1}, z_1 z_2] = (z_1^2 + z_2)L_{\underset{\sim}{g}}^{M-2} + z_1 L_{\underset{\sim}{g}}^{M-3} + L_{\underset{\sim}{g}}^{M-4},$$

with $\alpha, \beta \in \mathbb{R}$ so by the induction hypothesis and (4.3.2) we see that

$$(z_1^2 + z_2)L_{\underset{\sim}{g}}^{M-2} \in \Lambda. \tag{4.3.3}$$

But then

$$\begin{aligned}[L_{\underset{\sim}{g}}^2,[L_{\underset{\sim}{g}}^2,(z_2+z_1^2)L_{\underset{\sim}{g}}^{M-2}]] &= [L_{\underset{\sim}{g}}^2,[L_{\underset{\sim}{g}}^2, L_g^2,(z_2+z_1^2)]]L_{\underset{\sim}{g}}^{M-2} \\ &= 12\, L_{\underset{\sim}{g}}^M\end{aligned}$$

and so $L_{\underset{\sim}{g}}^M \in \Lambda,\ \forall\ M \geqslant 1$. This immediately implies that $\mathbb{R}[z_1,\ldots,z_{n-1}] \subset \Lambda$ since, $\forall\ |\alpha| \geqslant 0$

$$z_1^{\alpha_1} z_2^{\alpha_2} \cdots z_{n-1}^{\alpha_{n-1}} = \alpha_1! \cdots \alpha_{n-1}!\, \mathrm{ad}_{z_2}^{\alpha_1} \cdots\cdots \mathrm{ad}_{z_n}^{\alpha_{n-1}} (L_{\tilde{g}}^{|\alpha|})$$

so all these monomials are in Λ.

Now, from Th^{m} (4.2.3), $L_{\tilde{g}}$ is a surjective linear map from $\mathbb{R}[z_1,..,z_{n-1}]$ onto itself (this is actually a slight modification of Th^{m} (4.2.3) obtained by noticing that if ϕ is a polynomial in $z_1,..,z_{n-1}$, then $L_{\tilde{g}}(\phi) \triangleq \sum_{i=1}^{n} z_{i-1} \frac{\partial\phi}{\partial z_i} = \sum_{i=1}^{n-1} z_{i-1} \frac{\partial\phi}{\partial z_i}$ and $\sum_{i=1}^{n-1} z_{i-1} \frac{\partial}{\partial z_i}$ is shown to be a surjection on $\mathbb{R}[z_1,..,z_{n-1}]$.) Also we have

$$[L_{\tilde{g}}^2, \phi] = 2L_{\tilde{g}}(\phi)L_{\tilde{g}} + L_{\tilde{g}}^2(\Phi) \qquad \forall \phi \in \mathbb{R}[x_1,..,x_{n-1}]$$

Thus we see that

$$(4.3.4) \qquad \mathbb{R}[x_1,..,x_{n-1}]L_{\tilde{g}} \subset \Lambda.$$

We can now complete our proof of the claim (4.3.1) by considering the bracket of $L_{\tilde{g}}^2$ with z_n^2 which is readily seen to be

$$[L_{\tilde{g}}^2, z_n^2] = 2z_n z_{n-1} L_{\tilde{g}} + \mathrm{ad}_{L_{\tilde{g}}}^2 (z_n^2)$$

Since the second term on the R.H.S. is an element of Λ, it follows that $z_n z_{n-1} L_{\tilde{g}} \in \Lambda$. A simple induction using the brackets $[L_{\tilde{g}}, z_n z_{n-k} L_{\tilde{g}}]$ and (4.3.4) shows that $\{z_n z_{n-k} L_{\tilde{g}};\ 1 \leq k \leq n\} \subset \Lambda$, thus proving the claim (4.3.1).

[REMARK: We have used, without specific mention, the assumption that $n \geq 2$ in deriving (4.3.1), since (4.3.3) is invalid without this hypothesis, requiring as it does the existence of the coordinate z_2]

It remains to show that the polynomials in z_n with coefficients in $\mathbb{R}[z_1,..,z_{n-1}]$ are elements of Λ for which it is sufficient to prove that $z_n^m \phi \in \Lambda$ $\forall$ $m \geq 0$, and $\phi \in \mathbb{R}[z_1,...,z_{n-1}]$. For $m = 0,1$ this is easily seen to be the case using (4.3.1) since

$$[z_n L_{\tilde{g}}, \phi] = z_n L_{\tilde{g}}(\phi)$$

and $L_{\tilde{g}}$ is a surjection. Similarly, if $z_n^m \phi \in \Lambda$, then

$$[z_n L_{\tilde{g}}, z_n^m \phi] = z_n^{m+1} L_{\tilde{g}}(\phi) + m z_n^m z_{n-1} L_{\tilde{g}}(\phi)$$

so by the surjectivity of $L_{\tilde{g}}$ an inductive argument shows that $z_n^m \mathbb{R}[z_1,\ldots,z_{n-1}] \subset \Lambda \ \forall \ m \geq 0$. But

$$\mathbb{R}[z_1,\ldots,z_n] = \bigoplus_{m\geq 0} z_n^m \mathbb{R}[z_1,\ldots,z_{n-1}]$$

Completing the proof of the lemma.

□

So far we have only made limited use of the structure theory of graded polynomial systems developed in Chapters I and II. This situation is rectified in the following result which, although fundamental to the proof of Th^{m} (4.3.1), is of independent interest as no specific assumptions are made on the particular gradations involved.

THEOREM 4.3.3

Let $\Sigma = \{f,g,h\}$ be a minimal g.c.p.f. on $\mathbb{R}^n$ with $\{\mathbb{R}[x_1,\ldots,x_n], L_g^2\} \subset \Lambda(\Sigma)$ and L_g a surjective map from $\mathbb{R}[x_1,\ldots,x_n]$ onto itself. Then $\Lambda(\Sigma) = W_n$.

Proof

Since Σ is in g.p.f., the remarks in §4.1 mean that the generators of the estimation algebra take the form

$$F = -L_{\hat{f}} + \tfrac{1}{2} L_g^2 - \tfrac{1}{2}h^2 - \operatorname{div} \hat{f}\,, \qquad G = h.$$

But div $\hat{f}$ and $\frac{1}{2}h^2$ are polynomials so by hypothesis we see that $L_{\hat{f}} \in \Lambda(\Sigma)$. Further, if $\phi \in \mathbb{R}[x_1,\ldots,x_n]$ we find that

$$[F,\phi] = L_g(\phi)L_g + \psi$$

for some polynomial ψ. The surjectivity of L_g therefore implies that $\mathbb{R}[x_1,\ldots,x_n]\, L_g \subset \Lambda(\Sigma)$. Now, using this fact and a simple induction on the identity

$$[L_{\hat{f}}, \phi L_g] = L_{\hat{f}}(\phi)L_g + \phi[L_{\hat{f}}, L_g]$$

It is easily seen that $\mathbb{R}[x_1,\ldots,x_n]\operatorname{ad}^k_{L_{\hat{f}}} L_g \subset \Lambda(\Sigma)$. Consequently, using the algebraic identities of lemma (3.1.2) we obtain

(4.3.5) $$\mathbb{R}[x_1,\ldots,x_n] \otimes \mathscr{S}(\Sigma) \subset \Lambda(\Sigma)$$

However, we have assumed that Σ is in g.c.p.f. so that $\hat{\Sigma}$ is also in g.c.p.f. Hence, by Cor^y (1.3.4), $D_1(\mathbb{R}^n)$ (the space of all polynomial vector fields on $\mathbb{R}^n$) is contained in $\Lambda(\Sigma)$. In particular, this means that the set of generators $\tilde{\Lambda} = \{x_i^2,\ x_i^2 \frac{\partial}{\partial x_i},\ x_i x_{j+1};\ 1 \leq i \leq n,\ 1 \leq j \leq n-1\}$ is a subset of the estimation algebra. But Th^m (4.0.1) states that $\tilde{\Lambda} \cup \{\frac{\partial^2}{\partial x_i^2};\ 1 \leq i \leq n\}$ generates W_n, so to prove this theorem it remains to show that $\frac{\partial^2}{\partial x_i^2} \in \Lambda$ for $1 \leq i \leq n$. We achieve this by using the nilpotent strucure of $\mathscr{S}(\hat{\Sigma})$ to demonstrate that $\mathbb{R}[x_1,..,x_n] \otimes \mathscr{S}(\Sigma)^{\otimes 2} \subset \Lambda(\Sigma)$ and again appealing to Cor^y (1.3.4).

So let $\{\mathscr{S}^k;\ 1 \leq k \leq p\}$ denote the descending central series of $\mathscr{S}(\hat{\Sigma})$. Then $\forall \phi \in \mathbb{R}[x_1,..,x_n]$ and $X \in \mathscr{S}^p$ since $[L_g, X] = 0$ we must have

$$[L_g^2, \phi X] = 2L_g(\phi)L_g X \in \Lambda$$

and hence

(4.3.6) $$\mathbb{R}[x_1,\ldots,x_n]L_g X \subset \Lambda \qquad \forall X \in \mathscr{S}^p$$

Inductively, we assume that (4.3.6) holds $\forall X \in \mathscr{S}^{p-k}$ and $k = 0,..,K-1$. Then, if $Z \in \mathscr{S}^{p-K}$, an application of Lemma (3.1.2) yields

$$[L_g^2, \phi Z] = 2L_g(\phi)L_g Z + \phi\ \{2L_g[L_g, Z] - \mathrm{ad}_{L_g}^2 Z\}$$

Since $L_g \in \mathscr{S}^1$ it follows immediately that (4.3.6) is valid for $k = K$ and hence is true $\forall X \in \mathscr{S}(\Sigma)$.

Similarly,

$$[L_{\hat{f}}, \phi L_g X] = L_{\hat{f}}(\phi)L_g X + \phi[L_{\hat{f}}, L_g]X + \phi L_g[L_{\hat{f}}, X]$$

and since $\mathscr{S}(\hat{\Sigma})$ is $\mathrm{ad}_{\hat{f}}$-invariant (so $[L_{\hat{f}}, X] \in \mathscr{S}(\hat{\Sigma})$) (4.3.6) and an induction imply that

(4.3.7) $$\mathbb{R}[x_1,..,x_n]YX \subset \Lambda \qquad \forall Y \in \mathcal{R}^1(\hat{\Sigma}),\ X \in \mathscr{S}(\hat{\Sigma})$$

where $\mathcal{R}^1(\hat{\Sigma})$ is defined in Th^m (4.1.1). We assume now that (4.3.7) is true for all

$Y \varepsilon R^j(\hat{\Sigma})$, $1 \leqslant j \leqslant k-1$. Then $\forall\, Z \varepsilon R^1(\hat{\Sigma})$, $Y \varepsilon R^{k-1}(\hat{\Sigma})$, $X \varepsilon \mathscr{S}(\hat{\Sigma})$

$$[Z, \phi YX] = L_Z(\phi)YX + \phi[Z,Y]X + \phi Y[Z,X]$$

which by definition of $R^k(\hat{\Sigma})$ and the inductive hypothesis shows that (4.3.7) is also valid for $Y \varepsilon R^j(\hat{\Sigma})$ with $1 \leqslant j \leqslant k$, hence for all j. But

$$\mathscr{S}(\hat{\Sigma}) = R^1(\hat{\Sigma}) + R^2(\hat{\Sigma}) + \ldots + R^p(\hat{\Sigma})$$

It therefore follows that (4.3.7) is true $\forall\, X, Y \varepsilon \mathscr{S}(\hat{\Sigma})$, or $\mathbb{R}[x_1, \ldots, x_n] \otimes \mathscr{S}(\hat{\Sigma})^{\otimes 2} \subset \Lambda$. From Cor^y (1.3.4) we have now

$$\hat{D}_2(\mathbb{R}^n) = \sum_{k=0}^{2} \mathbb{R}[x_1, \ldots, x_n] \otimes \mathscr{S}(\hat{\Sigma})^{\otimes k} \subset \Lambda(\Sigma)$$

where $D_2(\mathbb{R}^n)$ is the space of all polynomial second order differential operators. Thus $\Lambda(\Sigma)$ contains $\{\frac{\partial^2}{\partial x_i^2};\ 1 \leqslant i \leqslant n\}$ as required.

$\square$

The proof of Theorem (4.3.1) is now a trivial consequence of Lemma (4.3.2) and the above result, since we have shown that the hypothesis of Th^m (4.3.1) imply that $L_{\tilde{g}}$ is surjective and both $\mathbb{R}[z_1, \ldots, z_n]$ and $L^2_{\tilde{g}}$ are elements of the estimation algebra. As we remarked in the introduction, this result is in a sense unsatisfactory as we have had to assume that $\{L_{\tilde{g}}, L^2_{\tilde{g}}, z_n^2\} \subset \Lambda$. The result remains of interest, however, since no explicit hypotheses, other than the requirement that it be in polynomial form, have been made on the structure of the drift vector field $\hat{f}$. There may be implicit restrictions on $\hat{f}$ needed to guarantee the existence of the above generators but these have yet to be determined.

We conclude this section and the thesis by applying Th^m (4.3.1) to the system inspiring the constructions of this chapter namely

$$\left\{\begin{array}{l} \left\{\begin{array}{ll} dx_1 & = dw \\ dx_2 & = x_1 dt + x_1 dw \end{array}\right. \\ \;\; dy \;\;\; = x_2 dt + dv \end{array}\right.$$

for which $L_{\hat{f}} = (x_1 - \frac{1}{2}) \frac{\partial}{\partial x_2}$, $L_g = \frac{\partial}{\partial x_1} + x_1 \frac{\partial}{\partial x_2}$ and $h(x) = x_2$. Thus,

$$F = \tfrac{1}{2}L_g^2 - L_{\hat{f}} - \tfrac{1}{2}x_2^2 - 1 \qquad \text{and } G = x_2$$

Since g is already in the d.i.o. form we have to show that $\{L_g, L_g^2, x_2^2\} \subset \Lambda$. Now using the identities of Chapter III, we find

$$[F,G] = x_1L_g + x_1 + \tfrac{1}{2} \triangleq X_1$$

$$[[F,G],G] = x_1^2 \triangleq X_2$$

$$[F,x_1^2] = 2x_1L_g + 1 \triangleq X_3$$

Comparing X_1 with X_3 we see that $x_1 \varepsilon\Lambda$. But

$$[F,x_1] = L_g$$

Thus, $L_g \varepsilon\Lambda$ and consequently $\mathbb{R} \subset \Lambda$ since $[L_g,x_1] = 1$ which in turn, from the form of X_3, shows that $x_1L_g \varepsilon\Lambda$. Now it is readily seen that $[L_g,L_{\hat{f}}] = \frac{\partial}{\partial x_2}$, so

$$[F,L_g] = -[L_{\hat{f}} + \tfrac{1}{2}x_2^2, L_g] = \frac{\partial}{\partial x_2} + x_1x_2 \triangleq X_4$$

and

$$[F,X_4] = -\tfrac{1}{2}[x_2^2, \frac{\partial}{\partial x_2}] + [F,x_1x_2]$$

$$\Rightarrow x_1x_2 + 2(x_2 + x_1^2)L_g \triangleq X_5 \ \varepsilon\ \Lambda$$

Moreover, since $[X_5,x_1^2] \Rightarrow x_1^3 + x_2x_1\varepsilon\Lambda$ and $[x_1L_g, x_1x_2 + x_1^3] \Rightarrow x_1x_2 + 4x_1^3\varepsilon\Lambda$ we obtain x_1^3 and $x_1x_2\ \varepsilon\Lambda$. In particular, we can bracket F with x_1^3 to obtain

$$[F,x_1^3] = 6x_1^2L_g + 3x_1 \triangleq X_6$$

But x_1 is already an element of the estimation algebra so by comparing X_6 with X_5 and using the previous comments we have now shown

$$\{\frac{\partial}{\partial x_2}, x_1^2L_g, x_2L_g, x_1^3, x_1x_2\} \subset \Lambda.$$

Now let us examine the expansions of $[F,x_2L_g]$ and $[F,[F,G]]$. We find

$$[F,x_2L_g] = x_1L_g^2 + x_2\frac{\partial}{\partial x_2} + x_1x_2^2 + x_1L_g + \tfrac{1}{2}L_g$$

$$\mathrm{ad}_F^2G = L_g^2 + x_1\frac{\partial}{\partial x_2} + x_1^2x_2 + L_g$$

implying that $X_7 \triangleq x_1 L_g^2 + x_2 \frac{\partial}{\partial x_2} + x_1 x_2^2$ and $L_g^2 + x_1^2 x_2 + x_1 \frac{\partial}{\partial x_2} \triangleq X_8$ are in Λ. Similarly

$$[X_7, L_g] = - L_g^2 - x_1 \frac{\partial}{\partial x_2} - \{x_2^2 + 2x_1^2 x_2\} \triangleq X_9.$$

Adding X_8 and X_9 gives the function $(x_2^2 + x_1^2 x_2) \varepsilon \Lambda$. However

$$[[F, x_1^3], x_1 x_2] = [3x_1^2 L_g, x_1 x_2] = 3x_1^2 x_2 + 3x_1^4$$

and $$[[F, x_1^3], x_1^3] = 3x_1^2 L_g, x_1^3 = 9x_1^4$$

Thus, $x_2^2 \varepsilon \Lambda$. Also, as an added bonus of these calculations, an examination of X_8 now reveals that $L_g^2 + x_1 \frac{\partial}{\partial x_2} \varepsilon \Lambda$. But

$$[L_g^2 + x_1 \frac{\partial}{\partial x_2}, x_1 L_g] = 2L_g^2 - x_1 \frac{\partial}{\partial x_2}$$

and, hence, $L_g^2 \varepsilon \Lambda$ completing the proof for this example.

BIBLIOGRAPHY

Abraham, R. & Marsden, J.E. [1]: *The Foundations of Mechanics* Benjamin/Cummings, Reading, Mass. (1978).

Beneš, V.E. [1]: "Exact Finite Dimensional Filters for Certain Diffusions with Nonlinear Drift", stochastics, 5, pp 65-92 (1981).

Brockett, R.W. [1]: "Volterra Series and Geometric Control", automatica, 12, pp 167-176, (1976).

[2]: "Some Remarks on Finite Dimensional Nonlinear Estimation", in 'Analyse des Systemes', Asterisque, vol. 75-76 pp 47-55, (1980).

[3]: "Classification and Equivalence in Estimation Theory". 1979 Conf. on Decision and Control, IEEE, N.Y. (1979).

Bucy, R.S. & Joseph, P.D. [1]: *Filtering For Stochastic Processes with Applications to Guidance*. Wiley Interscience, New York, (1968).

Conn, J.F. [1]: *Non Abelian Minimal Closed Ideals of Transitive Lie Algebras*. Math. Notes 25, Princeton University Press, (1981).

Crouch, P.E. [1]: "Dynamical Realisations of Finite Volterra Series". SIAM J. Cont. & Optimisation, 19 pp 177-202, (1981).

Cyrot-Normand, D. & Monaco, S. [1]: "The Immersion under Feedback of a Multidimensional Nonlinear System into a Linear System". Int. J. Cont. 38 pp 245-261 (1983).

Davis, M.H.A.

[1]: "Pathwise Nonlinear Filtering". In Hazewinkel & Willems [1].

[2]: "A New Approach to Filtering for Nonlinear Systems". Proc. I.E.E. 128D, pp 166-172 (1981).

Doss, H.

[1]: "Liens entre Équation Différentielles Stochastiques et Ordinaires". Ann. Inst. H. Poincare, 13 pp 99-125, (1977).

Duncan, T.E.

[1]: "Evaluation of Likelihood Functions". Inf. & Control, 13, pp 62-74 (1968).

Elliott, D.L.

[1]: "A Consequence of Controllability". Jnal. of Diff. Equns, 10 pp 364-370 (1971).

Flato, M., Simon, J., Snellman, H., & Sternheimer, D.

[1]: "Simple facts about Analytic Vectors and Integrability". Ann. Scient. Ec. Norm, Sup. 4e. serie, 5, pp 423-434 (1972).

Fliess, M.

[1]: "Realisations Locale des Systèmes Non-linéaires, algèbres de Lie Filtrées transitives et séries génératrices non-commutatives". Invent. Math., 71, pp 521-537, (1983).

[2]: "Un Outil Algebraique: Les Series Formelles Non Commutatives". In "Mathematical Systems Theory", Ed. G. Marchesini & S. Mitter, Lect. Notes Econom. & Math. Systems, 131, pp 122-148. Springer-Verlag, Berlin (1976).

[3]: "Realisations of Nonlinear Systems and Abstract Transitive Lie Algebras". Bull. Amer. Math. Soc. 2 pp 444-446, (1980).

[4]: "The Unobservability Ideal for Nonlinear Systems". IEEE tr. AC-26 pp 592-593, (1981).

Fliess, M. & Kupka, I. [1]: "A Finiteness Criterion for Nonlinear Input-Output Differential Systems". SIAM J. Cont. & Opt., 21, pp 721-728, (1983).

Fujisaki, M., Kallianpur, G. & Kunita, H. [1]: "Stochastic Differential Equations for the Nonlinear Filtering Problem". Osaka J. Math., 9, pp 19-40, (1972).

Gauthier, J.P. & Bornard, G. [1]: "Observability for any u(t) of a Class of Nonlinear Systems". IEEE tr. AC-26, pp 922-926, (1981).

Goodman, R.W. [1]: *Nilpotent Lie Groups* Lect. Notes. in Math. 562 Springer-Verlag, Berlin, (1976).

Guillemin, V. & Stenberg, S. [1]: "An Algebraic Model of Transitive Differential Geometry". Bull. Amer. Math. Soc. 70, pp 16-47, (1964).

Hermann, R. & Krener, A.J. [1]: "Nonlinear Controllability & Observability" IEEE tr. AC-22 pp 728-740, (1977).

Hazewinkel, M. & Willems, J.C. [1]: *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*. Proc. NATO-ASI, Les Arcs' 1980, Reidel, Holland, (1981).

Hazewinkel, M. & Marcus, S.I. [1]: "On Lie Algebras and Finite Dimensional Filtering". Stochastics 7, pp 29-62, (1982).

Hazewinkel, M., Marcus, S.I. & Krishnaprasad, P.S. [1]: "Current Algebras and the Identification Problem". Stochastics, 11, pp 65-101, (1983).

[2]: "System Identification and Nonlinear filtering: Lie Algebras". Proc. 20th IEEE Conf. on Decision & Control 1981 IEEE (1981).

Hazewinkel, M., Liu, C.H. & Marcus, S.I. [1]: "Some Examples of Lie Algebraic Structure in Nonlinear Estimation". In Proc. 1980 J.A.C.C. San Francisco, pp 978-983.

Hijab, O. [1]: "A Realisation Theory For Nonlinear Stochastic Systems". Proc. 22nd IEEE Conf. on Decis. & Control 1983. IEEE, (1983).

[2]: Minimum Energy Estimation Ph.d. Thesis, University of California, Berkely (1980).

Jacobson, N. [1]: Basic Algebra I. W.H. Freeman & Co., San Francisco (1974).

[2]: Lie Algebras. Dover, New York, (1979).

Jacubczyk, B. & Respondek, W. [1]: On Linearisation of Control Systems. Bull. Acad. Polon. Sci. Ser. Sci. Math., 28, pp 517-522 (1980).

Jazwinski, A.H. [1]: Stochastic Processes & Filtering Theory. Academic Press, New York, (1970).

Jorgensen, P.E.T. [1]: "Representations of Differential Operators on a Lie Group". Jnal. of Funct. Anal. 20, pp 105-135, (1975).

Kailath, T. [1]: "Some Topics in Linear Estimation". In Hazewinkel & Willems [1].

Krener, A.J. & Lesiak, C.M. [1]: "The Existence & Uniqueness of Volterra Series for Nonlinear Systems". IEEE tr. AC-23, pp 1090-1095, (1978).

Kallianpur, G. [1]: Stochastic Filtering Theory Appns. of Math. 13, Springer-Verlag, New York, (1980).

Kunita, H. [1]: "Stochastic Partial Differential Equations Connected with Nonlinear Filtering". In Nonlinear Filtering & Control, Ed. S.K. Mitter & A. Moro; Lect. Notes in Math. 972, Springer-Verlag, New York, (1982).

[2]: "On Backward Stochastic Differential Equations". Stochastics, 6, pp 293-313, (1982).

Kirillov, A.A. [1]: Elements of the Theory of Representations Springer-Verlag, New York (1976).

Levine, J. [1]: "Une Classe de Systèmes Nonlinéarires à Temps Continu admettant des Filtres de Dimension Finie". Preprint.

Liptser, R.S. & Shiryaev, A.N. [1]: Statistics of Random Processes, I Springer-Verlag, New York, (1977).

Marcus, S.I. & Davis, M.H.A. [1]: "An Introduction to Nonlinear Filtering". In Hazewinkel & Willems [1].

Marcus, S.I. & Willsley, A.S. [1]: "Algebraic Structure & Finite Dimensional Nonlinear Estimation". SIAM J. Math. Anal., 9, pp 312-327, (1978).

Mortensen, R.E. [1]: "Maximum Likelihood Recursive Nonlinear Filtering". J. Opt. Theory & App. 2, pp 386-394, (1968).

Michel, D. [1]: "Regularity of Conditional Laws in Nonlinear Filtering Theory and Stochastic Calculus of Variations". In Hazewinkel & Willems [1].

Moore, R.T. [1]: "Exponentiation of Operator Lie Algebras on Banach Spaces". Bull. Amer. Math. Soc. 71, pp 903-908, (1965).

Nijmeier, H. [1]: "Observability of a Class of Nonlinear Systems: a Geometric Approach". Richerche di Automatica, 12, pp 1-19, (1981).

Ocone, D. [1]: Topics in Nonlinear Filtering Ph.d. Thesis, M.I.T., (1980).

[2]: "Finite Dimensional Estimation Algebras in Nonlinear Filtering". In Hazewinkel & Willems [1].

Omori, H. & de la Harpe, P. [1]: "About Interactions between Banach-Lie Groups and Finite Dimensional Manifolds". J. Math. Kyoto Univ. 12, pp 543-570, (1972).

Omori, H. [1]: "On Banach-Lie Groups acting on Finite Dimensional Manifolds". Tôhoku Math. Journ. 30, pp 223-250, (1978).

Palais, R.S. [1]: "Natural Operations on Differential Forms". Trans. Amer. Math. Soc. 92 pp. 125-141, (1959).

Singer, I. & Sternberg, S. [1]: "The Infinite Groups of Lie and Cartan". Journal d'Analyse Math. 15, pp 1-115, (1965).

Spivak, M. [1]: A Comprehensive Introduction to Differential Geometry, I (2nd Ed.) Publish & Perish, Berkely, (1979).

Simon, J. [1]: "On the Integrability of Representations of Finite Dimensional Lie Algebras". Comm. Math. Phys. 28. pp 39-46, (1972).

Sussman, H.J.

[1]: "Existence and Uniqueness of Minimal Realisations of Nonlinear Systems". Math. Syst. Theory 10, pp 263-284, (1977).

[2]: "Orbits of Families of Vector Fields and Integrability of Distributions". Trans. Amer. Math. Soc., 180, pp 171-188, (1973).

[3]: "Semigroup Representations, Bilinear Approximation of Input-Output Maps and Generalised Inputs". In *Mathematical Systems Theory*, Ed. G. Marchesini and S. Mitter. Lect. Notes in Econom. and Math. Syst. Th. 131 pp 172-191, Springer-Verlag, Berlin (1976).

[4]: "On the gap between Deterministic and Stochastic Ordinary Differential Equations". The Ann. of Prob., 6, pp 19-41, (1978).

[5]: "Single Input Observability of Continuous Time Systems". Math. Syst. Theory, 12, pp 371-393, (1979).

Sussman, H.J. & Jurdjevic, V.

[1]: "Controllability of Nonlinear Systems". J. Diff. Equns., 12, pp 95-116, (1972).

Wei, J. & Norman, E.

[1]: "On the Global Representation of the Solutions of Linear Differential Equations as a Product of Exponentials". Proc. Amer. Math. Soc., 15, pp 327-334, (1964).

Wiener, N. [1]: *Extrapolation, Interpolation and Smoothing of Stationary Time Series* Technology Press & J. Wiley N.Y., 1949.

Wonham, W.M. [1]: "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering". J. SIAM, Series A: Control 2, pp 347-369, (1965).

Zakai, M. [1]: "On the Optimal Filtering of Diffusion Processes". Z. Wahr. verw. Geb. 11 pp 230-243, (1969).

Nijmeijer, H. [2]: "Feedback Decomposition of Nonlinear Control Systems". IEEE tr. AC-28,7, (1983).