ARTICLE IN PRESS

Automatica ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Survey paper

The early days of geometric nonlinear control☆

Roger Brockett

School of Engineering and Applied Sciences, Harvard University, United States

ARTICLE INFO

Article history:
Received 2 January 2014
Received in revised form
11 April 2014
Accepted 29 May 2014
Available online xxxx

Keywords:
Nonlinear control
Differential geometry
Differentiable manifold
Lie group
Bilinear systems
Volterra series
Vector fields
Lie brackets
Feedback linearization
Controllability
Carleman linearization
Maximum principle
Optimal control
Singular control
Stochastic differential equations
Hypoellipticity
Attitude control
Nonholonomic systems
Quantum control
Feedback stabilization

ABSTRACT

Around 1970 the study of nonlinear control systems took a sharp turn. In part, this was driven by the hope for a more inclusive theory which would be applicable to various newly emerging aerospace problems lying outside the scope of linear theory, and also by the gradual realization that tools from differential geometry, and Lie theory in particular, could be seen as providing a remarkably nice fit with what seemed to be needed for the wholesale extension of linear control theory into a nonlinear setting. This paper discusses an initial phase of the development of geometric nonlinear control, including material on the broader context from which it emerged. We limit our account to developments occurring up to the early 1980s, not because the field stopped developing at that point but rather to limit the scope of the project to something manageable. Even so, because of the volume and diversity of the literature we have had to be selective, even within the given time frame.

1. Introduction

This paper discusses an initial phase of the development of geometric nonlinear control, including material on the broader context from which it emerged. We limit our account to developments occurring up to the early 1980s, not because the field stopped developing at that point but rather to limit the scope of the project to something manageable. Even so, because of the volume and diversity of the literature we have had to be selective, in some cases only skimming the surface. The many applications, ranging from momentum wheel control of satellites and the generation of robotic trajectories, to the acrobatics of falling cats, provide effective advertisement for the relevance of the subject and some of these applications are discussed here as well. With the goal of reaching an audience wider than just those involved in this area of research, and mindful of the fact that some of the concepts involved are not everyday fare for engineers, we include considerable background material to enable suitably motivated readers not working in the area to better appreciate the what and why that lies behind the who and when.

Much of the early work in this area was related to applications in fields as disparate as satellite control, path planning for mobile robots and the design of excitation sequences for magnetic resonance spectroscopy. As was the case when state space methods were being introduced to describe linear systems, sometimes new methods are dismissed as being too theoretical, however in the case of differential geometric control this attitude seems to reflect an unfamiliarity with the mathematics being used rather than any

☆ This work was supported in part by the Army Research Office under contract number W911NF-12-1-0350. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Editor John Baillieul.

E-mail address: brockett@seas.harvard.edu.

http://dx.doi.org/10.1016/j.automatica.2014.06.010
0005-1098/© 2014 Elsevier Ltd. All rights reserved.

lack of applicability. In fact, without attempting to mount an argument involving ever changing, and methodologically questionable, citation statistics and impact factors, suffice it to say that when control science is viewed in a wider context, the additional scope these ideas give to the theory and practice is impressive. Section 12 of this survey sketches a few applications that lie quite outside the reach of linear theory. In general, we give more prominence here to those parts of the theory that relate to applied problems.

2. Differential equations and transfer functions

In the 1960s the development of state space methods recast the way students of control learned about linear systems, changing the subject from one based on Laplace transforms and transfer functions to one in which vector spaces and first order linear differential equations took the center of the stage. At that time linear algebra was not part of the standard curriculum for engineering students (Matlab did not exist!) and this state space revolution had the effect of making linear algebraic ideas part of the everyday vocabulary of control engineers. While this was beneficial in that it opened up additional points of contact with mathematics and physics, it had the side effect of putting the field of control somewhat apart from previously neighboring engineering disciplines. It even created something of a schism within the field itself, as some declined to learn the new language. However, driven by concomitant developments in optimization theory, notably the maximum principle of Pontryagin, Boltyanskii, Gamkrelidze, and Mishchenko (1959), numerical methods for trajectory optimization (Kelley, Kopp, & Moyer, 1966), the work of Fisk (1963), Itô (1946) and Stratonovich (1963) on stochastic differential equations, the Kalman–Bucy filter Kalman and Bucy (1961), etc., and aided by the availability of well written expositions such as Kalman's papers on linear systems (Kalman, 1960a, 1963) and some excellent text books on linear algebra (e.g. Gantmacher (1959) and Halmos (1958)) this "linear revolution" proceeded quickly, if not painlessly.[1] The main ideas underpinning linear system theory have a clear mathematical structure, and key concepts such as controllability and feedback invariants, provided a rough template for how a more comprehensive nonlinear theory might develop.

In the middle of the 19th century, the work of Airy (1840), on tracking telescopes and that of Maxwell (1868) on fly-ball governors, control was closely linked to differential equations. These, and other early applications of control technology, were usually concerned with physical systems, modeled in this way. One or two decades later, as the importance of technologies based on the transmission of power and information over electrical networks grew, there emerged a competing way to describe systems based on frequency response and transform methods. The "operational calculus" of Oliver Heaviside eventually led to a distinctively different, "systems" point of view, which often proved to be more practical for these new problems. Eventually these methods were firmly supported by the theory of the Laplace transform and, in time, led to effective ways of thinking about feedback and feedback compensation, in the process developing concepts now often associated with the names Carson, Black, Nyquist and Bode, etc. By the mid 1940s this approach was widely taught in electrical engineering departments.

A decade later the pendulum began to swing the other way. In the 1950s the influential group at RIAS, organized by Lefschetz and LaSalle, played a central role shifting work in America back to the earlier, differential equations dominated, point of view. The RIAS group popularized recent developments in the field of differential equations and control, bringing the considerable body of theory under development in the Soviet Union to the attention of a wider circle of engineers. Especially prominent in this regard were questions related to stability in the sense of Liapunov, including the focus on concrete problems such as the Lur'e problem Aizerman and Gantmcher (1963) and (Lur'e & Postnikov, 1944), relating specifically to nonlinear feedback. This was the Sputnik/Cold War era and developments in the Soviet Union were of great interest, particularly in the United States.[2] The link to technology via control theory helped to revitalize certain problem areas in differential equations and the combination of differential equation methods with frequency response ideas often proved to be remarkably effective in solving concrete problems and explaining their significance in engineering terms. Particularly noteworthy in this regard is the result of Popov–Kalman–Yakubovich on stability (Kalman, 1971; Popov, 1962; Yakubovich, 1962).

2.1. New ideas from differential geometry

Around 1970 the study of nonlinear control systems took a sharp turn. In part, this was driven by the hope for a more inclusive theory which would be applicable to various aerospace problems lying outside the scope of linear theory, and also by the gradual realization that tools from differential geometry, and Lie theory in particular, could be seen as providing a remarkably nice fit with what seemed to be needed for the wholesale extension of linear control theory into a nonlinear setting. For systems describable by differential equations, this geometric approach seemed to hold the promise of a *systematic* development of nonlinear control, something that had been completely missing in the past.[3] Problems such as finding conditions under which the describing function could be validated and understanding the behavior of systems with hysteresis feedback, which had loomed large a decade earlier, suddenly seemed less important in comparison with what could be envisioned with these new methods. However there were impediments. New vocabulary and background material had to be digested, and, in stark contrast to what is available today, the expository literature in the area was sparse and uneven. For this reason the early paper of Hermann (1963), couched as it was in the language of "distributions in the sense of Chevalley" and Jan Kučera's work (Kučera, 1966) using Lie groups, took some time to be appreciated.[4] Indeed, there was a steep learning curve for engineers who wished to follow and contribute to these developments.

The 1972 David Mayne and I set out to improve this state of affairs by organizing a conference at Imperial College that brought together about 100, mostly youngish (≤ 35) people, with the idea of teaching and learning about how control problems fit in with differential geometric ideas. The lectures ranged in emphasis from applications to abstract theory, touching on a variety of topics. The proceedings (Mayne & Brockett, 1973) represent a faithful account of the lectures but do not fully capture the excitement that went

[1] Although rather eclectic, Bellman's book on matrix theory (Bellman, 1955) deserves to be mentioned in this context because of its large number of interesting references and suggestions for further work.

[2] Even so, there were important lines of work that did not receive the attention they might have. For example, the body of work on nonholonomic systems discussed in book by Neimark and Fufaev (1972), with its strong geometric flavor and extensive references to the Soviet literature on mechanics, does not appear to have played much of a role.

[3] About this time the "geometrization" of classical mechanics, as popularized by the stylish and audacious book of Abraham (with Marsden) (1968), began to attract a larger following and this provided further inspiration. See Brockett (1977).

[4] It is somewhat surprising that, notwithstanding the highly influential position Lefschetz held as the algebraic geometer of his day, and the geometric flavor of his book on differential equations (Lefschetz, 1957), I have seen no evidence that he investigated the possibility of using Lie theoretic methods for controllability.

along with the meeting. From a control engineers prospective, a new set of mathematical ideas were being coupled with things they already knew about. From the point of view of those already accomplished in the mathematics being used, it was a source of problems, pre-digested and put in mathematical form, on which they could use to advantage what they already knew. Contacts were made and seeds for future work were sown. In contrast with the famous 1911 Solvay Conference on physics, this conference did not just consolidate known material or reinforce the positions of a pre-existing establishment, it more nearly represented a scientific free for all; the only thing Imperial about it was the name of the host institution. Best of all, it brought together a group of talented people who found the subject interesting and whose subsequent work helped sustain the field for the next several decades. It even inspired a 25th anniversary conference, held in London in 1997.

3. Manifolds and vector fields

As mentioned, the concepts and prior results which formed the point of departure for this kind of nonlinear control, center around vocabulary and facts mostly unfamiliar in engineering disciplines. These include differentiable manifold, tangent bundle, vector fields, distributions and Lie algebras, etc., all prerequisite for appreciating the theorems of Frobenius, Chow and Rashevskii, and all that followed along these lines. Much of the background material is standard differential geometry, known in something like its present form for a century or more. Even so, before the 1970s its use had been largely limited to the study of problems internal to mathematics itself with little impact on engineering. In his nice monograph on applications of differential geometry Flanders (1963) makes a valiant attempt to show engineering relevance but, in fact, most of the applications he discusses relate to questions in mathematics and basic electromagnetism.

3.1. Choosing coordinates: an example

Although nowadays the word manifold appears frequently in the engineering literature, it is often used informally. In mathematics, the idea of a differentiable manifold is all about specifying precise rules for choosing coordinates and describing which spaces can be given coordinates with "nice" properties. The idea is so central to this story that it makes sense to spend a few words on its development, starting with a widely studied example having a rich history and representing many points of view.

Consider the problem of controlling the orientation of a rigid body using torques generated by gas jets. The equations of motion for the angular velocities and the orientation of a rigid body with exogenous torques, have been known since the time of Euler. The equations for the angular velocities, ignoring the orientation of the body, are most elegantly expressed in body fixed coordinates. They are nonlinear but beautifully symmetric,

$$I_1\dot{\omega}_1 = (I_2 - I_3)\omega_2\omega_3 + u_1$$
$$I_2\dot{\omega}_2 = (I_3 - I_1)\omega_3\omega_1 + u_2$$
$$I_3\dot{\omega}_3 = (I_1 - I_2)\omega_1\omega_2 + u_3.$$

Here the u_i are the applied torques, also expressed in body fixed coordinates. This system, with the input torques restricted in various ways, provides variety of concrete examples to illustrate nonlinear controllability results (Brockett, 1976a). However, Euler's representation of the equations describing the *orientation* of the rigid body relative to a fixed frame of reference using the familiar (θ, ϕ, ψ) is less appealing. Although completely suitable for some problems, such as the precessing top, the introduction of Euler angles detracts from the natural symmetry and introduces singularities at points where the definition of the Euler angles are degenerate. As an alternative, consider expressing the orientation of the body using an orthogonal matrix Θ who is columns are the coordinates of a set of unit vectors representing an orthogonal triad fixed in the body and expressed in a laboratory frame. The evolution of Θ is then given by $\dot{\Theta} = \Omega\Theta$ where Ω is a skew-symmetric matrix formed from the angular velocities expressed in body fixed coordinates

$$\Omega = \begin{bmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{bmatrix}.$$

In these coordinates the equations of motion can be expressed succinctly and symmetrically,

$$\dot{\Omega} = [\hat{I}, \Omega\hat{I}^{-1}\Omega] + U; \qquad \dot{\Theta} = \Theta\Omega$$

where $\hat{I}$ is the inertial tensor of the body. The skew-symmetric matrix U accounts for the external torques, appropriately scaled by the inertial tensor. Here, and throughout, when dealing with matrices, $[A, B] = AB - BA$.

Although redundant, in the sense that the matrices Ω and Θ appear to require a total of 18 scalars for their specification, as compared with the corresponding description using Euler angles, these equations have a symmetry reflecting the physical symmetry, do not require that the inertial matrix be diagonal, and are even mathematically meaningful when Θ and Ω are of any positive dimension. Moreover, because $\Omega = -\Omega^T$, it is specified by three scalars and the Cayley transform allows one to parametrize orthogonal matrices as $\Theta = (I - S)(I + S)^{-1}$ with $S = -S^T$, provided that $I + \Theta$ is invertible. Thus six parameters suffice to specify Ω and Θ. A short calculation shows that $S = (I - \Theta)(I + \Theta)^{-1}$ and that S satisfies the differential equation

$$\dot{S} = -(1/2)(I - S)\Omega(I + S).$$

Thus, just as in the case of the Euler representation, we have a set of six scalars that parametrize the motion, provided that we exclude some "bad" points. In fact, there are many coordinate systems, such as those based on quaternions, that have found wide use in representing the equations of motion. The concept of a differentiable manifold is designed to clarify the relationships between the various possibilities and to provide a language and a body of facts which allow one to explain, for example, why any three parameter scheme for the representation of orientation, such as the Euler angles or the representation as $\Theta = (I-S)(I+S)^{-1}$, must have singularities and why any singularity free embeddings into a cartesian space, such as the (Ω, Θ) representation appearing here, invariably requires the description to lie in a higher dimensional space.

3.2. Manifolds

There are many important problems in nonlinear control that only involve "local behavior" valid in the neighborhood of a point or a specific trajectory. In such cases all the analysis can be thought of as taking place in a open subset of $\mathbb{R}^n$. Examples of this type will appear below and in discussing them no mention of the word manifold is necessary. However, even when doing this sort of local analysis there is often an implicit or explicit agreement on what changes of variable are to be permitted. Usually these are required to be differentiable throughout the domain of interest. For example, in studying asymptotic stability it is important to restrict attention to changes of coordinates that are differentiable at the equilibrium point. Although $\dot{x} = -x$ admits $x = 0$ as an exponentially stable solution, the change of coordinates $z = \log x$ gives $\dot{z} = -1$, an equation that does not even have an equilibrium point. At the same time, there are important problems such as those involved in the control of tumbling satellites, for which a global point of view is necessary. It is with these in mind that we provide the following short discussion, intended to make the reader feel comfortable with phrases such as, "the topological space X can be given the structure of an n-dimensional $\mathcal{C}^k$ manifold".

The concept of a manifold, or rather the various concepts of what one might mean by a manifold, goes back to Riemann's 1854 lecture (Riemann, 1868) describing higher dimensional geometry and, more explicitly, to the 1895 paper of Poincaré (1895) on topology. Poincaré's approach involves seeing manifolds as subsets of $\mathbb{R}^n$ described by imposing smooth constraints. Consider a set of constraints of the form $\phi_1(x) = 0, \phi_2(x) = 0, \ldots, \phi_k(x) = 0$ where $\phi_i : \mathbb{R}^n \to \mathbb{R}$ and the matrix of partial derivatives

$$M = \begin{bmatrix} \partial\phi_1/\partial x_1 & \partial\phi_1/\partial x_2 & \cdots & \partial\phi_1/\partial x_n \\ \partial\phi_2/\partial x_1 & \partial\phi_2/\partial x_2 & \cdots & \partial\phi_2/\partial x_n \\ \cdots & \cdots & \cdots & \cdots \\ \partial\phi_k/\partial x_1 & \partial\phi_k/\partial x_2 & \cdots & \partial\phi_k/\partial x_n \end{bmatrix}$$

is of rank $n - m$, both on the set $X = \{x|\phi_i(x) = 0; i = 1, 2, \ldots, k\}$ and in a neighborhood of X. In this case X admits the structure of an m-dimensional manifold. For example, $X = \{x|x_1^2+x_2^2+x_3^2-1 = 0\}$ defines a two-dimensional manifold. In adopting this point of view one allows the use of coordinates to describe points in X that are related to the familiar coordinate system in $\mathbb{R}^m$ with varying degrees of smoothness. For example, one might require different coordinate systems to be related by a r times differentiable change of coordinates and in this way get a $\mathcal{C}^r$ manifold. If a vector field is to define a flow on a manifold $X = \{x|\phi_i(x) = 0, i = 1, 2, \ldots, k\}$ then it is necessary and sufficient that

$$\frac{d\phi_i}{dt} = \left\langle \frac{\partial\phi_i}{\partial x}, f(x) \right\rangle = 0; \quad i = 1, 2, \ldots, k.$$

Today there are many readable accounts of this material but around 1970 the mainstays were Auslander and MacKenzie (1963), Bishop and Richard (1964) and Sternberg (1964).

In some settings the requirement that a manifold should necessarily be a subset of $\mathbb{R}^n$ seems artificial so a second "intrinsic" definition emerged. It starts with the idea of a *locally euclidean space* and attaches coordinate charts with compatibility conditions requiring that when the charts overlap the respective coordinate descriptions should be related by a transformation having a specific order of differentiability. It was only with the work of Hassler Whitney in the 1930s that it was proven that any n-dimensional manifold described in this more intrinsic way could be embedded in $\mathbb{R}^m$ for some sufficiently large m and so the intrinsic definition did not really provide additional generality. In summary, the words "X admits the structure of a n-dimensional differentiable manifold" just means that for each point in X there is a neighborhood of the point that can be mapped in a one-to-one and onto way to a neighborhood of the origin in $\mathbb{R}^n$ and that there exists an agreed upon set of coordinates with which to make such an identification. A differentiable manifold is not just a set of points with a topology; the definition also includes restrictions on the coordinates that can be used for identifying points in the "pre-manifold" with points in $\mathbb{R}^n$.

Of course $\mathbb{R}^n$ with its usual topology and usual coordinates is the most common manifold but there is no shortage of other useful objects to which we can attach coordinates in such a way as to give them the structure of a manifold. To start with, any open subset of $\mathbb{R}^n$, such as $\{x|x \neq 0\}$, can be given the structure of a manifold. Other examples include, n-dimensional spheres and n-by-n orthogonal matrices. The set of orthogonal matrices is an example of a particularly interesting and "applications friendly" class of manifolds known as group manifolds, or Lie groups. Special cases will play a role in the later sections. An example of a set that does not have the structure of a manifold in a nontrivial way is the subset of $\mathbb{R}^2$ consisting of the points $\{(x, y)|xy = 0\}$. This set cannot be given the structure of a one-dimensional manifold because it is not locally euclidean at $x = y = 0$. Even though everywhere else it can be mapped to the real line in one-to-one and onto way, this is not possible in a neighborhood of the origin.[5]

3.3. Vector fields

In most of the literature relating to dynamical systems, if the description of some situation is given in terms of n simultaneous first order differential equations $\dot{x} = f(x)$, one says that $f(x)$ defines a vector field. If called on to elaborate, one might suggest that it be visualized as a field of arrows whose direction and magnitude at a particular value of x is given by the vector $f(x)$. However in differential geometry the description of a vector field is often expressed differently. The data in f are associated with the first order differential operator $F = \sum f_i(x)\partial/(\partial x_i)$ and it is this operator that is called a vector field. These two ideas are connected by the fact that if $\psi(x)$ is a scalar valued, differentiable function of x and if x satisfies $\dot{x} = f(x)$ then

$$\frac{d}{dt}\psi(x) = \left(\sum_{i=1}^{n} f_i(x)\frac{\partial}{\partial x_i}\right)\psi(x)$$

or simply $\dot{\psi} = F\psi$ where F is the first order differentiable operator appearing on the right-hand side of the equation. In general the practice of calling the same thing by two different names and/or using two different notations for the same thing has little to recommend it but in this case there are compelling reasons. Both

$$f(x) = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \quad \text{and} \quad F = f_1\frac{\partial}{\partial x_1} + f_2\frac{\partial}{\partial x_2} + \cdots + f_n\frac{\partial}{\partial x_n}$$

find use.

3.4. An emerging vision

Returning now to nonlinear control, we can say in retrospect that around 1970 the field was in the first stages of a wholesale extension of linear theory. This program sought to provide a generalization of linear theory along the following lines.

CONCEPT	LINEAR CASE	NONLINEAR CASE
state space	$\mathbb{R}^n$	differentiable manifold X
description	$\dot{x} = Ax + Bu$	$\dot{x} = f(x, u)$
coordinate changes	linear transformations	diffeomorphisms
subdomains	subspaces of $\mathbb{R}^n$	sub manifolds $\hat{X} \subset X$
controllability	$[B, AB, \ldots, A^{n-1}B]$	Lie brackets
feedback laws	$u \mapsto Mu + Kx$	$u \mapsto M(x)u + k(x)$

4. Bilinear models and Lie algebras

In the extension of linear theory into nonlinear domains, the study of bilinear systems forms a useful link. They include linear systems as a special case and the model fits a number of the most successful applications of nonlinear theory. In addition, their

[5] Milnor (1956) showed, for example, that a seven-dimensional sphere, as a topological space with the usual topology, can be given the structure of a seven-dimensional differentiable manifold in multiple, inequivalent, ways, thus illustrating the need for caution when ascribing the structure of a differentiable manifold to a space for which only a definition of continuity is given a priori.

Please cite this article in press as: Brockett, R., The early days of geometric nonlinear control. Automatica (2014), http://dx.doi.org/10.1016/j.automatica.2014.06.010

study provides an introduction to Lie algebras, Lie groups, and homogeneous spaces, introducing relevant algebraic structures while sidestepping some technical issues associated with more general situations. This class of systems is defined as those representable as

$$\dot{x} = Ax + \sum_{i=1}^{m} u_i B_i x; \qquad y = Cx; \quad x \in \mathbb{R}^n.$$

Their analysis, both local and global aspects, is relatively free of technicalities. Following the 1968 paper of Rink and Mohler (1962), two University of Rome groups, namely, (Bruni, Di Pillo, & Koch, 1971, 1974; D'Allesandro, Isidori, & Ruberti, 1972), developed detailed theories about the structure of these systems, exploring the similarities and differences with linear systems.

In the 1930s and 40s Norbert Weiner and Y. W. Lee and their students at MIT, had written extensively about the use of Volterra expansions in engineering problems such as the evaluation of the distortion caused by nonlinear elements in amplifier circuits. In this work, the usual starting point was a block diagram description of the system based on operators. Rather little attention was paid to differential equation descriptions of systems. Thus it marked a new direction when it was observed in Bruni et al. (1974) that bilinear systems have convergent Volterra series whose kernels are relatively easy to compute and characterize (D'Allesandro et al., 1972). In fact, the Peano–Baker series, expressing the solution of $\dot{x}(t) = A(t)x(t)$ in terms of a convergent power series,

$$x(t) = \left(I + \int_0^t A(\sigma_1)\, d\sigma_1 + \int_0^t \int_0^{\sigma_1} A(\sigma_1)A(\sigma_2)\, d\sigma_2 d\sigma_1 + \cdots \right) x(0)$$

provides the means for finding the Volterra series expansion for the solution of $\dot{x} = Ax + \sum u_i B_i x$. To explain this briefly, consider the special case in which A and B_i are time-invariant and u is a scalar. Introduce $z = e^{-At}x$ satisfying the equation

$$\dot{z} = \sum u_i e^{-At} B_i e^{At} z; \qquad z(0) = x(0).$$

The first three terms in the Peano–Baker series are then

$$I + \int_0^t e^{-A\sigma_1} B e^{A\sigma_1} u(\sigma_1)\, d\sigma_1 + \int_0^t \int_0^{\sigma_1} e^{-A\sigma_1} B e^{A(\sigma_1-\sigma_2)} B e^{A\sigma_2} u(\sigma_1)u(\sigma_2)\, d\sigma_2 d\sigma_1$$

and the rest follow similarly. The corresponding terms in the Volterra series expansion for x are obtained by premultiplying by e^{At} and post multiplying by $x(0)$. If u is bounded the series is convergent on any finite interval. The kernel functions are causal and if A and B_i are constant with $Ax(0) = 0$, the kernels are invariant under time translation, $t \mapsto t + t_0$.

4.1. Structure theory of bilinear systems

Structure theories for the constant coefficient case, i.e., A, B_i, and C all constant, were developed by a number of authors (Brockett, 1972a; Bruni et al., 1971; D'Allesandro et al., 1972; Rink & Mohler, 1962). The results parallel previous developments for linear systems but with significant differences. Given a time invariant system of the form

$$\dot{x} = Ax + \sum_{i=1}^{m} u_i B_i x; \qquad y = Cx; \qquad x(0) = x_0 \in \mathbb{R}^n$$

distinguish between the situation in which the reachable set from the given initial condition lies in a proper subspace of $\mathbb{R}^n$ and when it does not. In the former case, there exists a linear transformation $x \mapsto Tx$ such that A and B_i are all block upper triangular and, along with $Tx(0)$ take the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}; \qquad B = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}; \qquad x(0) = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}.$$

Instead of focusing on concepts of controllability/reachability defined in terms of the reachable set (see the next subsection), the structure theory works with the smallest linear space that contains the reachable set, acknowledging that it depends on the initial condition. In this way it is possible to reduce the system to an *irreducible system* with the same input–output behavior such that the reachable set from x_0 contains a basis for this linear space.[6] This does not mean that the reachable set for the reduced system is the entire vector space. As for observability, an initial state x_1 is indistinguishable from x_2 if for all u the difference between the trajectories lie in the kernel of C. This happens if and only if there is a subspace, invariant for A and each B_i, lying in the kernel of C. The papers (Brockett, 1972a; D'Allesandro et al., 1972) prove that for irreducibility defined in this way two irreducible bilinear systems defining the same input–output map are related by a similarity transformation, just as in the linear case.

Example. Consider the scalar input, $n + 1$-dimensional bilinear system having initial condition $x^T(0) = [1, 0, \dots, 0]$ and equation of evolution

$$\frac{d}{dt}\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} + u \begin{bmatrix} 0 & 0 & \cdots & 0 \\ b_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ b_n & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

If $y = Cx$ this system simulates the system $\dot{z} = Az + bu$; $z(0) = 0$; $y = Cz$ and is irreducible in the above sense if and only if the linear system is controllable.

4.2. Matrix Lie algebras and linear differential equations

A *Lie algebra* is a vector space together with a bilinear map, written as $[\cdot, \cdot]$, which maps $\mathcal{L} \times \mathcal{L}$ to $\mathcal{L}$. It is subject to two conditions, skew-symmetry, $[L_1, L_2] = -[L_2, L_1]$ and the Jacobi identity, $[L_1, [L_2, L_3]] + [L_2, [L_3, L_1]] + [L_3, [L_1, L_2]] = 0$. The elements of a Lie algebra form a linear space and to describe the Lie algebraic structure it is enough to identify a basis for the space, say $\{g_1, g_2, \dots, g_l\}$ and to specify the three index array Γ whose entries, γ^i_{jk}, are such that $[g_i, g_j] = \sum \gamma^k_{ij} g_k$. These are called the *structure constants* relative to the given basis. However, another way to characterize a Lie algebra is to construct a one-to-one correspondence between the abstract L_i and a set of matrices $\{T_i\}$ such that if L_i corresponds to T_i and if $[L_i, L_j] = \sum \gamma^k_{ij} L_k$ then we have $T_i T_j - T_j T_i = \sum \gamma^k_{ij} T_k$. In this case the T_i are said to provide a *representation* of the Lie algebra.

The 1972 paper (Brockett, 1973a) discusses various ways that Lie theory can be useful in the study of control problems, focusing on bilinear models. Useful as the Peano–Baker series is, there are other representations of the fundamental solution of $\dot{x}(t) = A(t)x(t)$ which provide further details about the structure of the

[6] The word "irreducible" is used here, rather than "minimal" to be consistent with the terminology of Schur's lemma, as used in the theory of group representations, and to avoid conflict with they way the word minimal is otherwise used in nonlinear control.

Please cite this article in press as: Brockett, R., The early days of geometric nonlinear control. Automatica (2014), http://dx.doi.org/10.1016/j.automatica.2014.06.010

solution. Suppose that $A(t)$ takes the form $A(t) = \sum \phi_i(t)A_i$ with the $A_1, A_2, \ldots, A_m$ constant matrices and the ϕ_i scalar functions of time. Now consider the possibility of finding a set of matrices $L_1, L_2, \ldots, L_k$ such that it is possible to express the fundamental solution of $\dot{x} = A(t)x$ as

$$\Phi(t) = e^{L_1 g_1(t)} e^{L_2 g_2(t)} \cdots e^{L_k g_k(t)}.$$

The time derivative of this product of exponentials can be written as

$$\frac{d}{dt}\Phi = \left(\dot{g}_1 L_1 + \dot{g}_2 e^{g_1 L_1} L_2 e^{-g_1 L_1} + \dot{g}_3 e^{g_1 L_1} e^{g_2 L_2} L_3 e^{-g_2 L_2} e^{-g_1 L_1} + \cdots\right)\Phi.$$

The identity

$$e^A B e^{-A} = B + [A, B] + \frac{1}{2!}[A, [A, B]] + \frac{1}{3!}[A, [A, [A, B]]] + \cdots$$

follows from power series expansion of the exponential. Using this repeatedly, it is not too difficult to prove that if $\{L_i\}$ form a basis for a Lie algebra that contains the A_i then it will be possible to find a differential equation for the g_i whose solution is such that the fundamental solution of $\dot{x} = A(t)x$ can be expressed in the form given. The differential equation for the g_i is nonlinear, and solutions may have a finite escape time, but the equations for the g_i will always be solvable for small values of $|t|$.

Contributors to this circle of ideas include (Chen, 1957, 1962; Magnus, 1954; Wei & Norman, 1963, 1964). One important aspect of this work is that it relates the intrinsic complexity of finding a solution of the linear equation $\dot{x} = \sum \phi_i(t)A_i x$ to the dimension of the lowest dimensional Lie algebra having $\{A_i\}$ in its linear span. This has proved to be important in understanding controllability, the sufficient statistics problem in nonlinear filtering, and various aspects of quantum control.

Given a Lie algebra $\mathcal{L}$, either each element of $\mathcal{L}$ can be expressed in one or more ways as a bracket involving other elements of $\mathcal{L}$ or else elements of the form $[\mathcal{L}, \mathcal{L}]$ form a proper sub algebra of $\mathcal{L}$. More generally, associated with any Lie algebra $\mathcal{L}$ there is a chain of sub algebras defined inductively by $\mathcal{L}_0 = \mathcal{L}$ and $\mathcal{L}_i = [\mathcal{L}, \mathcal{L}_{i-1}]$. This is called the *lower central series of* $\mathcal{L}$ and if the containment is proper at each stage, terminating with $\mathcal{L}_f = 0$, then the Lie algebra is said to be *nilpotent*. The definition of a solvable Lie algebra is similar but differs in that the successor in the series is defined to be $\mathcal{L}_{i+1} = [\mathcal{L}_i, \mathcal{L}_i]$. If this series terminates in 0 then the algebra is said to be *solvable*. All nilpotent algebras are solvable but not conversely. For example, the set of all lower (or upper) triangular matrices form a solvable algebra whereas the set of all strictly lower (or upper) triangular matrices form a nilpotent Lie algebra. If $\dot{x} = \sum \phi(t)A_i x$ and the $\{A_i\}$ generate a nilpotent Lie algebra then the Peano–Baker series has only a finite number of terms. If the Lie algebra generated by the A_i is solvable then the fundamental solution can be expressed in terms of sums of integrals of matrix exponentials.

4.3. Bilinear systems and Lie groups

A matrix Lie group is a set of nonsingular matrices that is closed under multiplication and inversion and admits the structure of a differentiable manifold. Examples include the set of all nonsingular n-by-n matrices for a fixed integer n and the set of all n-by-n orthogonal matrices for a fixed integer n. Matrix Lie algebras are, in a sense, the logarithms of matrix Lie groups. That is, given $\mathcal{L}$, the set of all products of the form $e^{L_1}e^{L_2}\cdots e^{L_k}$ with $L_i \in \mathcal{L}$ is closed under multiplication and inversion. The dimension of the manifold defined by this set of products equals the number of linearly independent elements in $\mathcal{L}$. For example, the n-by-n orthogonal matrices are exponentials of skew-symmetric matrices and admit the structure of a Lie group of dimension $n(n-1)/2$. These "group manifolds" were studied extensively beginning in the latter part of the 19th century by the Norwegian mathematician Sophus Lie and are named after him. More generally, Lie groups are groups whose points can be given the structure of a differentiable manifold with the operations of multiplication and inversion being continuous. There are many textbooks devoted to these ideas; Rossmann (2003) is a recent example. The subject lies at the crossroads of algebraic group theory and differential geometry and is an integral part of physics and mathematics.

In the theory of first order linear differential equations $\dot{x} = Ax$ the fundamental solution is a matrix Φ which satisfies $\dot{\Phi} = A\Phi$ with $\Phi(0) = I$. The solution of the vector equation is then $\Phi(t)x(0)$. Of course Φ belongs to $Gl(n)$, the Lie group of n-by-n nonsingular matrices and $x(0)$ is an arbitrary point in $\mathbb{R}^n$. In the language of differential geometry, it is said that the group of all nonsingular matrices, written $Gl(n)$, *acts* on $\mathbb{R}^n$ sending x into Φx. More generally, a subgroup $\mathcal{G}_1 \subset \mathcal{G}l(n)$ is said to act on a set $S \subset \mathbb{R}^n$ if for all $s \in S$ and all $G \in \mathcal{G}_1$ $Gs \in S$. If, in addition, for any two points $s_1, s_2 \in S$ there exists $G \in \mathcal{G}_1$ such that $Gs_1 = s_2$ then $\mathcal{G}_1$ is said to *act transitively* on S and S is said to be a *homogeneous space*. For example, the n-by-n orthogonal matrices act transitively on the set of all unit vectors in $\mathbb{R}^n$, making the unit sphere a homogeneous space.

In studying the reachable set for bilinear systems, the relevant Lie algebra consists of the smallest linearly closed set of matrices that contain A and $B_1, B_2, \ldots, B_m$ while being closed under the bracket operation introduced above. Apparently the first papers specifically targeting reachability problems for bilinear systems are those of Kučera (1966, 1967) who investigated the evolution of an invertible matrix satisfying $\dot{X} = \left(A + \sum B_i\right)X$. Brockett (1972b) investigated reachability problems on Lie groups and made contact with earlier work on linear and bilinear systems via the idea of a homogeneous space. This paper also explored some special properties associated with solvable and compact groups and gave a Lie algebraic theorem on controllability with a drift term strong enough to cover linear systems. For the latter, consider a bilinear system $\dot{x} = Ax + \sum u_i B_i x$ under the assumption that $[ad_A^k(B_i), B_j] = 0$ for all i and j and $k = 1, 2, \ldots$. (Here, and below, $ad_A(\cdot) = [A, \cdot]$, $ad_A^2(\cdot) = [A, [A, \cdot]]$, etc.) In this situation the effect of the drift term, Ax, and the effects of the individual u_i are separable in the sense that x can be expressed as

$$x(t) = e^{At}\Phi_1(u_1)\Phi_2(u_2)\cdots\Phi_m(u_m)x(0)$$

where Φ_i is of the form e^{H_i} with H_i in the Lie algebra generated by $ad_A^k(B_i)$. This fact was used to show that the effect of the drift can be accounted for explicitly and that the reachable set at time t is the set of all x of the form $x = e^{At}e^H$ with H an arbitrary element of the Lie algebra generated by $ad_A^k(B_i)$. This result was subsequently generalized by Hirschorn (1973) using properties of special classes of Lie algebras which play an important role in several aspects of the subject. Reference (Brockett, 1972b) also considers nonlinear observability. The paper (Brockett, 1973b) makes use of properties of compact groups to develop a theory of systems evolving on spheres; this work is directly applicable to controllability issues in quantum control.

5. Linear analytic models

Assuming that $f(x, u)$ is has the required differentiability with respect to u at $u = 0$, a Taylor series expansion gives

$$\dot{x} = f(x, u) = f_0(x) + uf_1(x) + \cdots.$$

The 1963 papers by Hermes and Haynes (1963) and Hermann (1963) restrict their attention to systems in "linear analytic form", i.e.,

$$\dot{x} = f(x) + \sum_{i=1}^{m} u_i g_i(x)$$

with f and g_i being analytic in x. Subsequently this model has been used extensively in the control literature both as a starting point for controllability studies and also in the study of canonical forms for systems. On one hand, this seems to represent a considerable generalization with respect to the bilinear model but, as will be discussed later, bilinear systems have a remarkable ability to approximate systems of this more general form. In fact, although the techniques used appear to be rather different, the successes and failures in the study of linear analytic systems are not that different from those encountered in the bilinear situation.

5.1. Vector fields and Lie brackets

Before discussing various advances that were made in the control of linear analytic systems it is necessary to define what is meant by the Lie bracket as an operation mapping a pair of vector fields into a third. Initially, it is sufficient to consider vector fields defined in the neighborhood of a point in $\mathbb{R}^n$ because the definition of the Lie bracket is entirely local. Under mild assumptions, an autonomous differential equation $\dot{x} = f(x)$ determines, at least for small values of $|t|$, a unique integral curve passing through a given point. If $\dot{x} = f(x)$ then $\ddot{x} = f_x f$ where f_x is the Jacobian of f. Thus the first three terms in the Taylor series expansion of $x(t)$ gives

$$x(t) \approx x_0 + f(x_0)t + f_x(x_0)f(x_0)t^2/2.$$

Now, suppose that there are two differential equations defined on the same space, $\dot{x} = f(x)$ and $\dot{x} = g(x)$. If x satisfies $\dot{x} = f(x)$ on the interval $0 \le t < \epsilon$ and satisfies $\dot{x} = g(x)$ for $\epsilon \le t \le 2\epsilon$ we can estimate $x(2\epsilon)$ as follows. (All calculations are correct to second order only.) Because the initial value for the second part of the path is not x_0 but rather, $x_0 + f\epsilon + f_x f\epsilon^2/2$, we have

$$x(2\epsilon) \approx x_0 + \epsilon f + f_x f\epsilon^2/2 + g(x_0 + f\epsilon)\epsilon + g' g\epsilon^2/2$$

again, accurate to second order. The Taylor series expansion of $g(x_0 + f\epsilon)$ gives $g(x_0 + f\epsilon)\epsilon \approx g(x_0)t + g_x f\epsilon^2 + g_x g\epsilon^2/2$, still correct to second order, and finally,

$$x(2\epsilon) = x_0 + \epsilon(f + g) + f_x f\epsilon^2/2 + g_x f\epsilon^2 + g_x g\epsilon^2/2$$

where f, f_x, g, g_x are all to be evaluated at x_0. Now consider $x(2\epsilon)$ to be the half-way point along a path that is to be completed by following $\dot{x} = -f(x)$ for ϵ units of time and then following $\dot{x} = -g(x)$ for an additional ϵ units of time. It will be seen that in evaluating the end point of the complete path, the first order terms cancel out but the second order terms do not. Working through the details gives

$$x(4\epsilon) - x_0 = (g_x f - f_x g)\epsilon^2 \stackrel{\text{def}}{=} [f, g]\epsilon^2.$$

For f and g differentiable, the quantity $[f, g]$ is called the Lie bracket of f and g. Clearly, $[f, g] = -[g, f]$, and assuming that f, g, h are twice differentiable, a further calculation shows that the Jacobi identity, $[f, [g, h]] + [g, [h, f]] + [h, [f, g]] = 0$ holds.

Remark on notation. Observe that in terms of $F = \sum f_i \frac{\partial}{\partial x_i}$ and $G = \sum g_i \frac{\partial}{\partial x}$, the bracket $[f, g]$ corresponds to $FG - GF$. However, in the case of linear vector fields written as Ax and Bx, there is a conflict involving the definition of matrix multiplication on one hand and composition of linear differential operators on the other. This leads to $[Ax, Bx] = (BA - AB)x = [B, A]x$.

5.2. Lie algebras, distributions and controllability

From its definition, it is clear that the Lie bracket should play a role in the theory of the controllability of systems of the form $\dot{x} = f(x)u_1 + g(x)u_2$. The four segment path described above suggests that not only is it possible to steer x in the directions defined by f and g but also to create displacements in the direction $[f, g]$. Moreover, iterating on this idea, it seems that it should be possible to generate displacements in the directions $[f, [f, g]]$, $[g, [f, g]]$, $[f, [f, [f, g]]]$ etc. As explained below, a precise statement along these lines, together with a proof was given in the papers of Chow (1939) and Rashevskii (1938) (independently). The relevance of their result, with some extensions, is the message contained in Hermann (1963) and elaborated on by Haynes and Hermes (1970). The proof uses the concept of a distribution along with the integrability theorem of Frobenius. In the case of linear control systems, the Cayley–Hamiltonian theorem can be invoked to show that the exploration of the Lie brackets has a definite stopping point. Baillieul (1981) showed that something similar is true for polynomial vector fields.

Along with an n-dimensional manifold X comes a $2n$-dimensional manifold $T(X)$ called the *tangent bundle of* X. Its formal definition makes precise the intuitive idea of a tangent plane being attached to X at each point so that in a small neighborhood N of a point in X the tangent bundle looks like a cartesian product $N \times \mathbb{R}^n$. We can interpret the vector f appearing in $\dot{x} = f(x)$ as defining a pair $(x, f(x))$ belonging to the tangent bundle. Likewise, given two differential equations $\dot{x} = f(x)$ and $\dot{x} = g(x)$, with $f(x) \ne g(x)$ the set $(x, \alpha f + \beta g)$ for α and β arbitrary real numbers, defines a two-dimensional subspace of the tangent plane at each point x. Instead of attaching the entire tangent plane at each point of X to get $T(X)$, a set of vector fields on X can be thought of as defining a subspace of the tangent plane at each point. The resulting structure is called a *distribution*[7] it is an assignment of a subspace of the tangent space which depends on the point in the manifold in a smooth way.

A distribution, defined as the span of a set of vector fields, $\{f_1, f_2, \dots, f_m\}$ on an n-dimensional manifold X, is said to have constant rank if the rank of the n-by-m matrix $[f_1, f_2, \dots, f_m]$ does not depend on x. The distribution is said to be *involutive* if any two vector fields taking on values in the distribution has the property that their Lie bracket also takes on values in this distribution. That is, the value of $[f_i, f_j]$ at each point x takes on values in the distribution and must not introduce a new direction. The theorem of Frobenius asserts that there is (at least locally) a sub manifold of X, say X_D, associated with an involutive distribution $\mathcal{D}$.[8] The theorem can be thought of as a generalization of the observation that if $f(x_0) \ne 0$ then there exists a change of variable, $z = \phi(x)$ defined in a neighborhood of x_0 such that the differential equation takes the form $\dot{z} = c$ with c constant.[9] Thus the solution of the differential equation, expressed in these coordinates, is simply $z(t) = z_0 + f(z_0)t$ and consequently there are $n - 1$ constants of motion.

In summary, what Chow and Rashevskii added is the statement that if we have a set of vector fields, $g_1, g_2, \dots, g_k$ which is not involutive, and if we form their involutive closure by adding to this collection the possible brackets $[g_i, g_j]$, brackets of brackets $[g_i[g_j, g_k]]$ etc., increasing the dimension of the distribution until it becomes involutive, then given any two points in the manifold whose existence is assured by the theorem of Frobenius, there is

[7] Sometimes called a "distribution in the sense of Chevalley" to distinguish it from the distributions in the sense of Laurent Schwartz which play a role in functional analysis.

[8] Although Frobenius proved this result, he was not the first to do so. A. Clebsch and F. Deahna working independently and on separate aspects, had already done the equivalent considerably earlier. Even so, we are following convention and referring to it as the theorem of Frobenius.

[9] The idea of the proof is as follows. Because $f(x_0) \ne 0$ it is nonzero throughout a neighborhood of x_0. The solution of the differential equation $\dot{x} = f(x)$ starting at x_1, near x_0, takes the form $x(t) = x_1 + f(x_1)t + h(t)$ where h is second order in t. Clearly it is possible to make a change of variables in (t, x) space introducing t as a coordinate and giving the equation the desired form.

a path which joins them and which can be generated by following solutions of just the original vector fields $g_1, g_2, \ldots, g_k$. Obviously this can be interpreted as a controllability result and the title of Rashevskii's paper suggests exactly that. Thus, to determine the reachable set for the system $\dot{x} = \sum g_i(x)u_i$, bracket the vector fields g_i and their brackets, etc. until further bracketing produces only vector fields which lie in a distribution $\mathcal{D}$. Then determine a corresponding Frobenius manifold which contains x_0. Given any second point x_1 on that manifold, and a positive number t_1 there exists a control $u(t)$ defined on $[0, t_i]$ which steers this system from x_0 to x_1. All this was reviewed, spelling the relevant assumptions, in Lobry's survey (Lobry, 1973).

5.3. Reachability with a drift term

This Frobenius–Chow–Rashevskii–Hermann sequence gives a nice result on controllability. However almost all real systems have a drift term and this severely limits its applicability. In fact, this line of thought does not even let one treat linear systems of the form $\dot{x} = Ax+Bu$. Exploiting the fact that the vector fields of a linear system form a solvable lie algebra, there is a result in Brockett (1972b) that covers linear systems from the Lie algebraic point of view but for the more general situation involving drift, Krener (1974) and, independently, Sussmann and Jurdjevic (1972) established a general result applicable to systems with drift. What they showed is that if one, in effect, treats the drift as if it were a controlled vector field and computes the Frobenius–Chow–Rashevskii manifold containing x_0, then the set of reachable points from x_0 contains some open set in this manifold. Without further assumptions, little more can be said.

Earlier on, in their 1967 book on optimal control (Markus & Ernest, 1967) Markus and Lee included a number of interesting asides, including a reference to Caratheodory in the context of controllability. More relevant here, they show that the system $\dot{x} = f(x, u)$ with $f(0, 0) = 0$ and f differentiable in both x and u is such that any point in a neighborhood of the origin can be reached from $x = 0$ provided that the linear system $\dot{x} = Ax + Bu$ is controllable where $A = (\partial f)/\partial x)$ and $B = (\partial f)/\partial u)$, both evaluated at $(x, u) = (0, 0)$. A proof can be made using the standard construction of u for the linearized system $\dot{x} = Ax + Bu$ together with an appeal to the inverse function theorem. Interpreted in the present context, this says that for

$$\dot{x} = f_0(x) + uf_1(x) + u^2 f_2(x) + \cdots$$

with $f(0) = 0$, the Lie brackets at $x = 0$ are

$$f_1(0); \qquad [f_0, f_1] = -\frac{\partial f_0}{\partial x} f_1\bigg|_0; \qquad [f_0, [f_0, f_1]] = \left(\frac{\partial f_0}{\partial x}\right)^2 f_1\bigg|_0 \cdots.$$

That is, if the linearized system is controllable then brackets of the form $ad_f^i(g_j)$ span and no brackets involving two or more g_i are necessary to generate a spanning set.

Local controllability problems: Assuming a system of the form $\dot{x} = f(x) + \sum u_i g_i$ with $f(0) = 0$ and u taking on values in $\mathbb{R}^m$, find necessary and sufficient conditions on f and g_i such that:

a. there exist a neighborhood N of $x = 0$ such that every point in N can be reached.
b. there exist a neighborhood N of $x = 0$ such that every point in N can be reached in arbitrarily small time.
c. there exist neighborhoods N_1 and $N_2 \subset N_1$ of $x = 0$ such that every point in N_2 can be reached in arbitrarily small time following a trajectory that does not leave N_1.

Such problems were widely studied in the 1970s and a variety of sufficient conditions were given, but general necessary and sufficient conditions proved elusive. In studying the reachable set, it proved to be useful to look at the way in which various distributions are built up, starting from f and $g_1, g_2, \ldots, g_m$. In particular, it is useful to organize the Lie brackets in terms of the fewest number of occurrences of the g_is needed to span a given sub distribution. See Sussmann (1978). From this point of view, an ordering in which brackets of the form $ad_f^k(g)$ (i.e., terms with g_i entering linearly) come before brackets with any number of fs and two g_i such as $[g_i, ad_f^k(g_j)]$ and then three g_i, etc. Just as the function $y = x^n$ maps $\mathbb{R}$ onto $\mathbb{R}$ if n is odd but not if it is even, examples suggest that for a scalar input system $\dot{x} = f(x) + ug(x)$ the effect of the drift term can be neutralized by brackets that are linear in the g_i but not necessarily those involving "squares", etc.

Example. Consider $\dot{x} = f(x) + ug$ in the form

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -x_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ x_1^p \end{bmatrix} u$$

Lie algebra, calculations continuing from

$$[g, f] = \begin{bmatrix} -1 \\ px^p \end{bmatrix}; \qquad ad_g^2(f) = \begin{bmatrix} 0 \\ \alpha_2 x_1^{p-1} \end{bmatrix}; \qquad ad_g^3(f) = \begin{bmatrix} 0 \\ \alpha_3 x^{p-2} \end{bmatrix}$$

show that $p + 1$ occurrences of g are necessary to get a set of brackets that span when evaluated at $x = 0$. Proceeding directly, by solving the first equation for u gives $\dot{x}_2 = x_1^p(\dot{x}_1+x_1)$. Consequently, for trajectories starting at the origin

$$x_2(t) = (p + 1)^{-1} x_1^{p+1}(t) + \int_0^t x_1^{p+1}(\sigma)\, d\sigma.$$

Thus if p is odd then $x_2(t) \geq 0$ and there are points in any neighborhood of the origin that are not reachable. If p is even let u be some control defined on $[0, \tau]$ having average value zero and pth moment nonzero. Such a control steers $(0, 0)$ to a point of the form $(0, a)$ and by scaling u we see that any point of this form can be reached. Because the equation for x_1 is controllable, given any b there is a control that steers $(0, 0)$ to (b, c). This shows that points of the form $(b, c + a)$ are reachable from $(0, 0)$ with a arbitrary. Thus if p is even, an odd number of occurrences of g suffices and any point is reachable from the origin.

If the drift term is such that the free motion is periodic for all initial conditions or, more generally, if it has some kind of "near periodicity" then the drift term is approximately reversible in the sense that if the free motion is periodic of period T then the solution at time $T - \epsilon$ is the same the solution of $\dot{x} = -f(x)$ at time ϵ. A result of this type was given by Lobry (1974), who postulated Poisson stability of a symmetric drift term on a compact manifold. Brockett (1976a) proved controllability under the hypothesis that the drift vector field generates an almost periodic solution for all initial conditions. The application of such ideas to various mechanical systems has been widely investigated.

Controllability properties along a reference trajectory have been investigated by Hermes (1974, 1976). Let $x(t)$ denote a trajectory of a control system with $x(0) = x_0$. Hermes defines the system to be locally controllable along a given trajectory if at time $t_1 > 0$ all points in some open neighborhood of $x(t_1)$ can be reached using solutions starting at x_0 and he developed sufficient conditions to determine local controllability along a reference trajectory. The papers of Hermes (1974, 1976, 1978) and Krener (1975) investigated local controllability via nilpotent approximations.

6. Systems and canonical forms

Even though much of the theory of linear systems focuses on coordinate free concepts such as controllability, eigenvalue placement, etc., special choices of coordinates, such as those leading to Jordan normal form, also play an important role. Likewise, in studying nonlinear systems special choices of coordinates are often useful. However, now there are a number of possibilities to be considered. One would certainly want to consider two systems as being equivalent if they are related by a diffeomorphism. But, the set of diffeomorphisms is infinite dimensional in any reasonable sense and difficult to characterize in a concrete way. It seems more promising to focus on quantities that are invariant under changes of coordinates. Such quantities, if they exist, might allow one to check whether or not two seemingly different control systems are the same to within a change of coordinates, and if not, what are the obstructions to finding a suitable coordinate change.

6.1. A global isomorphism theorem

A central idea in the theory of linear time invariant models is the fact that if $\dot{x} = Ax + Bu$; $y = Cx$ and $\dot{z} = Fz + Gu$; $y = Hx$ are both controllable, observable, and finite dimensional, and if they define the same input–output map, (that is, $Ce^{At}B = He^{Ft}G$) then there exists a nonsingular matrix T such that $TAT^{-1} = F$, $TB = G$ and $CT^{-1} = H$. That is, under the given circumstances, a representation of a linear time invariant system is unique to within a choice of basis.

The papers of Sussmann (1973, 1977) considers input–output models of the form

$$\dot{x} = f(x, u); \qquad y = \phi(x); \qquad u(t) \in U$$

where x takes on values in a manifold X and y in a manifold N. The question addressed, and we paraphrase here, is this. If there is a second system, with the same input set and out space,

$$\dot{z} = a(z, u); \qquad y = \psi(z); \qquad u(t) \in U$$

such that for each initial state $x(0)$ of the x-system there is an initial state $z(0)$ of the z-system such that the input–output maps of the two systems are the same, and for each initial state $z(0)$ there is a corresponding state $x(0)$ such that the input–output relation of the two maps are the same, then what additional conditions on the two systems are needed if one is to conclude that the systems are isomorphic in the sense that there exists a diffeomorphism ϕ such that $z = \phi(x)$.

As might be anticipated, it is necessary to assume that the vector fields are complete in the sense that for any admissible choice of u, the integral curves exist for all time. If the manifolds and the vector fields are analytic, if the systems both have the accessibility property, and if any two initial states are distinguishable, then the conclusion follows and there is a similar result in the $\mathcal{C}^\infty$ case if the vector fields have a certain symmetry. See Sussmann (1977) for more details and references to important preliminary material. This subject was further examined by Jakubczyk (1980).

6.2. Local equivalences

As noted above, if a smooth vector field is nonzero at a point then there exists a change of coordinates in the neighborhood of that point such that the vector field is constant. In this sense, in the neighborhood of a point, all smooth, nonzero vector fields are the same. However, in the neighborhood of a point where a vector field vanishes, things are more interesting. Near the end of the nineteenth century Poincaré and Liapunov (independently) considered the following question. If $\dot{x} = f(x)$ with f analytic and $f(0) = 0$, when is it possible to find a diffeomorphism ϕ such that $z = \phi(x)$ satisfies $\dot{z} = Az$ for some constant matrix A? That is, when is it possible to find a change of variables that recasts a nonlinear differential equation with an equilibrium point in linear form. They gave sufficient conditions for the existence of a ϕ, defined for x in some neighborhood of the origin, such that under the change of variables $z = \phi(x)$ the differential equation $\dot{x} = f(x)$ becomes $\dot{z} = f_\phi(z) = \phi_x f(\phi^{-1}(z))$ where

$$\phi_x = \left.\frac{\partial \phi}{\partial x}\right|_{x=\phi^{-1}(z)}$$

and $f_\phi(\cdot)$ is linear. The conditions involve only the eigenvalues of the Jacobian of f evaluated at zero. They require a strong form of non resonance; no integer combination of the eigenvalues of the Jacobian f_x can vanish, and they require that all the eigenvalues of the Jacobian should lie to one side of a line through the origin in the complex plane.

There is one very important property of Lie algebras of vector fields relating to change of coordinates. Let f_ϕ denote the transformed version of f given by $f_\phi(\cdot) = ((\partial\phi)/\partial x)f(\phi^{-1}(\cdot))$. If $[f, g] = h$ then $[f_\phi, g_\phi] = h_\phi$. That is to say, one can bracket and then change coordinates or one can change coordinates and then bracket and the result will be the same. In this way, a diffeomorphism ϕ defines a Lie algebra homomorphism. That this is so can be reasoned by reflecting on the fact that we get the same result if we solve a differential equation and then change coordinates or change coordinates and then solve the differential equation. This sometimes expressed by saying that the Lie bracket is a *natural operation*. One consequence of this is that if a set of vector fields, say $\{g_1, g_2, \ldots, g_m\}$ is an involutive collection and if ϕ is a diffeomorphism sending them into $\{\tilde{g}_1, \tilde{g}_2, \ldots, \tilde{g}_m\}$ then this distribution is also involutive.

6.3. Linear systems from a nonlinear point of view

It is sometimes convenient to avoid the discussion of time varying vector fields by adding an additional component to the state vector satisfying $\dot{x}_0 = 1$. In this way a time varying system in $\mathbb{R}^n$, $\dot{x} = \sum b_i(t)u_i$ is replaced by a time invariant system $\dot{x} = \sum b_i(x_0)u_i$; $\dot{x}_0 = 1$ evolving in $\mathbb{R}^{n+1}$. Krener (1973) considered the question of how to determine if there exists a change of coordinates such that a linear analytic system is transformed into a linear system of the form $\dot{x} = \sum b_i(t)u_i$.

A short calculation then shows that the vector fields

$$\tilde{b}_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}; \qquad \tilde{b}_1 = \begin{bmatrix} 0 \\ b_{11}(x_0) \\ \vdots \\ b_{1n}(x_0) \end{bmatrix}; \qquad \tilde{b}_2 = \begin{bmatrix} 0 \\ b_{21}(x_0) \\ \vdots \\ b_{2n}(x_0) \end{bmatrix}; \text{ etc.}$$

satisfy the commutation relations

$$[\tilde{b}_0, \tilde{b}_i] = \begin{bmatrix} 0 \\ \dfrac{\partial b_{i1}}{\partial x_0} \\ \vdots \\ \dfrac{\partial b_{in}}{\partial x_0} \end{bmatrix}; \qquad [ad^k_{b_0}(b_i), ad^l_{b_0}(b_j)] = 0; \quad 0 \le k, l \le n.$$

The central idea involved in investigating the existence of a change of variables that converts $\dot{x} = f(x) + \sum g_i(x)u_i$ to linear form is the fact that a change of variables defines a mapping from the Lie algebra associated with f and g_i to the Lie algebra associated with their transformed versions and this mapping is a Lie algebra homomorphism. Thus if there exists a change of variables that puts $\dot{x} = f(x) + \sum g_i(x)u_i$ in the linear form defined above, it is necessary that the Lie algebra generated by f and the g_i satisfy the commutation relations $[ad^k_f(g_i), ad^l_f(g_j)] = 0$. This condition turn out

Please cite this article in press as: Brockett, R., The early days of geometric nonlinear control. Automatica (2014), http://dx.doi.org/10.1016/j.automatica.2014.06.010

to be sufficient as well.[10] Notice that this solves the problem of determining the circumstances under which there exists a time dependent transformation $z=\phi(x,t)$ sending $\dot{x}=f(x)+\sum g_i(x)u_i$ to $\dot{z}=\sum b_i(t)u_i$. This does not solve the problem of finding conditions under which there exists a diffeomorphism $z=\phi(x)$ that transforms $\dot{x}=f(x)+\sum g_i(x)u_i$ to $\dot{z}=Az+\sum b_iu_i$. An answer to this latter question would subsume the Liapunov–Poincaré problem discussed above.

6.4. Embedding and representations

To set the stage for what is to come next, consider the pair of scalar input systems

$$\dot{x}=ax^2+2bx+c+u \quad \text{and} \quad \begin{bmatrix}\dot{z}_1\\ \dot{z}_2\end{bmatrix}=\begin{bmatrix}b & a\\ c & -b\end{bmatrix}\begin{bmatrix}z_1\\ z_2\end{bmatrix}+\begin{bmatrix}0\\ z_1\end{bmatrix}u.$$

It follows from the equation for z that z_1/z_2 satisfies

$$\frac{d}{dt}\left(\frac{z_1}{z_2}\right)=a\left(\frac{z_1}{z_2}\right)^2+2b\left(\frac{z_1}{z_2}\right)+c+u$$

and thus, as long as $z_2\neq 0$ the nonlinear system describing x is simulated by a bilinear system describing z. However, clearly there is no "change of variable" $y=\phi(x)$ such that the equation for y is bilinear. This illustrates the importance of distinguishing between what can be achieved by a diffeomorphism $\phi:X\to Z$ and what can be achieved by an embedding of a system in a higher dimensional space.

In Krener (1975) assumes a linear analytic model of the form $\dot{x}=f(x)+\sum g_i(x)u_i$ but now considers, as a target, the bilinear system $\dot{z}=Az+\sum u_iB_iz$. To simply ask if there is a diffeomorphism $z=\phi(x)$ such that z satisfies a bilinear equation would, again, run up against the Liapunov–Poincaré problem. Instead, Krener asks if in the neighborhood of a point x_0 there is a bilinear matrix system $\dot{Z}=AZ+\sum u_iB_iZ$ and a map ϕ such that $x=\phi(Z)$ for all u that is sufficiently small in an appropriate sense. Reasoning as above and using the fact the matrix Lie algebra generated by $A,B_1,B_2,\dots,B_m$ is necessarily finite dimensional, one sees that it is necessary that f and $g_1,g_2,\dots,g_m$ generate a finite dimensional Lie algebra. A theorem of Ado dating from 1935 (Ado, 1935) states that every finite dimensional Lie algebra has a matrix representation, and Krener uses this to show that the necessary condition is also sufficient.

7. Feedback invariants and feedback linearization

Given that ordinarily some form of feedback will be used to modify the performance of a system that is to be controlled, and given that the form of the feedback is often arbitrary except for bounds on the controls, the characterization of a system it is often more meaningful to focus on properties of the system that are invariant with respect to feedback. For example, unless the inputs are bounded, the reachable set of states is invariant under feedback so that in this case reachability questions depend only on properties that are feedback invariant.

[10] The Jacobi identity can be used to show that the condition $[ad_f^k(g_i),ad_f^l(g_j)]=0$ for all $k+l\leq r$ is equivalent to the condition that $[ad_f^s(g_i),g_j]=0$ equals zero for all $s\leq r$.

7.1. Linear systems with linear feedback

In 1970 Brunovsky (1970) and Rosenbrock (1970) published results describing the feedback invariants for a controllable linear system $\dot{x}=Ax+Bu$, assuming linear time invariant feedback and linear changes of variable in state and input. More explicitly, they considered transformations whereby x is replaced by $z=Px$, and u is replaced by $v=Mu+Kx$. This change results in the new description, $\dot{z}=(PAP^{-1}-M^{-1}KP^{-1})z+PBM^{-1}v$. The situation is captured by the block lower triangular transformation

$$\begin{bmatrix}z\\ v\end{bmatrix}=\begin{bmatrix}P & 0\\ K & M\end{bmatrix}\begin{bmatrix}x\\ u\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}x\\ u\end{bmatrix}=\begin{bmatrix}P^{-1} & 0\\ -M^{-1}KP^{-1} & M^{-1}\end{bmatrix}\begin{bmatrix}z\\ v\end{bmatrix}$$

where P and M are invertible matrices and K is arbitrary, as described by Wonham and Morse (1972). This lower triangular representation gives a concise formulation of the *linear feedback group*. Brunovsky and Rosenbrock showed, in rather different ways, that the only invariants with respect to change of basis and linear feedback are a certain set of positive integers (controllability indices) and that any controllable, n-dimensional system such that the rank of B equals the dimension of u can, through this type of transformation, be recast as a set of rank B decoupled systems whose size is fixed by these indices.

These indices can be related to the pair (A,B) in various ways, but for our purposes it is most convenient to introduce $r_k=\text{rank}\,[B,AB,\dots,A^{k-1}B]$. These integers are invariant under the action of the feedback group. From one point of view, the situation is summed up by saying that any linear time invariant system is feedback equivalent to a decoupled set of higher order systems $(d^i/dt^i=x^{(i)})$

$$x_1^{(k_1)}=u_1;\qquad x_2^{(k_2)}=u_2;\cdots;x_m^{(k_m)}=u_m,$$

and the k_i are determined by the r_i.

7.2. Nonlinear equivalence and feedback invariants

The problems raised in the NASA report by Meyer and Cicolani (1975) initiated a study of nonlinear feedback invariants. In extending linear theory to linear analytic systems, the first step is to identify an appropriate nonlinear version of a group action analogous to the group action used in the linear case. The definition should be such that the transformations take linear analytic systems into linear analytic systems and should include $x\mapsto\phi(x)$ with ϕ a diffeomorphism. The fact that the inputs are required to enter the equations linearly implies that only linear changes of variable are allowable for the inputs. Consider replacing u by $v=M(x)(u-k(x))$ with $M(x)$ nonsingular and analytic in x and k also analytic in x and then expressing the equations in terms of $z=\phi(x)$ with ϕ a diffeomorphism. In this way we get a family of systems which can be regarded as being equivalent. This definition of the feedback group which emerged out of the work of Brockett (1978) and Jakubczyk and Respondek (1980), is a substantial generalization of the linear case.

It is not to be expected that it will be possible to find a complete set of invariants for linear analytic systems unless further assumptions are made. Because the Lie algebra generated by f and $g_1,g_2,\dots,g_m$ is not feedback invariant (even in the linear case) one of the main tools used above, i.e. the fact that $x\mapsto\phi(x)$ induces a Lie algebra homomorphism, is ineffective here. However, if we restrict attention to a neighborhood of an equilibrium point, there exist invariants analogous to the controllability indices. If α and β are smooth scalar functions then

$$[f+\alpha g_i,\beta g_j]=\beta[f,g_j]+\left\langle\frac{\partial\beta}{\partial x},f\right\rangle g_j+\alpha\beta[g_i,g_j] -\beta\left\langle\frac{\partial\alpha}{\partial x},g_j\right\rangle g_i+\alpha\left\langle\frac{\partial\beta}{\partial x},g_i\right\rangle g_j.$$

This shows that if $[g_i, g_j] = 0$ then feedback alters the span of the brackets in the series $ad_f^k(g_i), ad_f^{k-1}(g_i), \dots, g_i$ a very limited way.

Making this more systematic, consider the way the Lie algebra of vector fields generated by f and $g_1, g_2, \dots, g_m$ builds up from its generators. First consider only brackets that are linear in the g_i. Such brackets generate a family of distributions, denoted by $\mathcal{D}_l$, and defined as

$$\mathcal{D}_0 = \text{span}\{g_1, g_2, \dots, g_m\}$$
$$\mathcal{D}_1 = \text{span}\{\mathcal{D}_0, [f, g_1], [f, g_2], \dots, [f, g_m]\}$$
$$\dots$$
$$\mathcal{D}_k = \text{span}\{\mathcal{D}_{k-1}, ad_f^k(g_1), ad_f^k(g_2), \dots, ad_f^k(g_k)\}.$$

We will say that these distributions are *regular* if the dimension of the space they span is the same at each point in a neighborhood of the origin and from now on, we limit attention to systems for which these distributions are regular. As described in Brockett (1978) it follows from the above identity that these distributions are feedback invariant. For $i = 0, 1, \dots, n-1$ the dimension of $\mathcal{D}_i$, is a feedback invariant for the system.

From this we see that if two systems whose drift terms vanish at $x = 0$ are feedback equivalent with a diffeomorphism that maps 0 to 0 then their corresponding $\mathcal{D}$ series must have the same dimension at each level and must have the given regularity properties.

Example. Consider the distinction between the system on the left and the one on the right.

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}x_2\\ f(x_1)\end{bmatrix} + \begin{bmatrix}0\\ 1\end{bmatrix}u; \qquad \begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}x_2^3\\ f(x_1)\end{bmatrix} + \begin{bmatrix}0\\ 1\end{bmatrix}u$$

both being two-dimensional, scalar input systems. For these systems $\mathcal{D}_0$ and $\mathcal{D}_1$ are spanned by

$$\begin{matrix}\text{left}\\ \text{system}\end{matrix}\begin{bmatrix}0\\ 1\end{bmatrix} \text{ and } \left\{\begin{bmatrix}0\\ 1\end{bmatrix}, \begin{bmatrix}1\\ 0\end{bmatrix}\right\} \begin{matrix}\text{right}\\ \text{system}\end{matrix}\begin{bmatrix}0\\ 1\end{bmatrix} \text{ and } \left\{\begin{bmatrix}0\\ 1\end{bmatrix}, \begin{bmatrix}3x^2\\ 0\end{bmatrix}\right\}.$$

These systems are not feedback equivalent with any diffeomorphism ϕ that maps 0 to 0 because at $x = 0$ the distributions are not of the same dimension.

7.3. Feedback linearization

Any linear subspace of the tangent bundle of $\mathbb{R}^n$ is integrable but, typically distributions are not and this is an important part of what makes the study of nonlinear feedback invariants more involved than what is found in the linear case. The space spanned by the columns of B, not only defines a subspace of the tangent bundle of $\mathbb{R}^n$ but it also defines a subspace of $\mathbb{R}^n$ itself which makes it possible to define a sub manifold whose tangent space is the range space of B. More generally, the same is true of the span of the columns of $[B, AB, \dots, A^{k-1}B]$. This already shows that if $\dot{x} = f(x) + \sum u_i g_i(x)$ is feedback equivalent to a controllable linear system then these distributions must be integrable.

From what was said above it is clear that a necessary condition for $\dot{x} = f(x) + \sum u_i g_i(x)$ with $f(0) = 0$ to be locally feedback equivalent to a controllable linear system with a diffeomorphism that maps 0 to 0 is that the distributions $\mathcal{D}_i$, identified above, must be integrable and that one of the distributions should satisfy dim $\mathcal{D}_k =$ dim x. They are also sufficient (Brockett, 1978; Jakubczyk & Respondek, 1980). There are various ways to prove this, one being organized around the idea of showing that these conditions imply that there is a function of $h(x)$ such that if $x \in \mathbb{R}^n$ then the first $n-1$ time derivatives do not depend on u. The original proof in Brockett (1978) is, in a certain sense, constructive and we sketch it here. For simplicity consider the scalar input case. The process proceeds by using feedback and changes of variable to create a set of intermediate systems of the form

$$\begin{bmatrix}\dot{x}_a\\ \dot{x}_b\end{bmatrix} = \begin{bmatrix}f(x_a) + Bx_b\\ Cx_b\end{bmatrix} + \begin{bmatrix}g_b(x_a)\\ e_n\end{bmatrix}u.$$

The idea is to successively apply feedback and changes of variables with the goal of pushing the nonlinear terms further and further away from the input. To advance from a form in which the dimension of the nonlinear part, x_a, is k to one where it is of dimension $k-1$ is to make a change of variables that makes g_b constant. Then use the integrability condition to show that x_k must enter f_a linearly and proceed.[11] One way to express the final result is to say: The system $\dot{x} = f(x) + ug(x)$ with equilibrium point at $x = 0$, is, in a neighborhood of 0, feedback equivalent to one of the form $\dot{x} = Ax + bu$ with (A, B) a controllable pair if and only if $\{ad_f^k(g)\}_{k=1}^{n-1}$ spans $\mathbb{R}^n$ at $x = 0$ and for k, m integers between zero and $n-1$ there exist d_i so that

$$[ad_f^k(g), ad_f^m(g)] = \sum_{i=1}^{\max(k,m)} d_i ad_f^{i-1}(g).$$

By the mid 1980s, related work on disturbance rejection (Hirschorn, 1981; Isidori, Krener, Claudio, & Monico, 1981), observer design through output injection (Krener, 1983), (f, g)-invariant subspaces (Byrnes & Krener, 1983; Hirschorn, 1981; Isidori, Krener, Claudio, & Monico, 1981) and decoupling (Freund, 1975; Hirschorn, 1981; Isidori et al., 1981) brought the theory of linear analytic systems to a level of completion quite comparable to that of multivariable linear system theory.

8. Stability and feedback stabilization

The period between 1950 and 1980 was an exceptionally fruitful one for the theory of differential equations. It saw the emergence of new results on structural stability, chaotic behavior, the famous KAM theory for Hamiltonian systems and a resurgence of interest in differential equations of celestial mechanics, both for their own sake and because of their relevance for aerospace engineering. In the 1950s work in the USSR elevated to prominence the earlier contributions of Liapunov. By incorporating new developments in the theory of dynamical systems, researchers were able to better understand nonlinear stability questions, including both the local behavior near a degenerate equilibrium point (Hahn, 1967; Malkin, 1959) and some global questions (Zubov, 1964). Important developments in Liapunov theory over this period include the work of Kurzweil (1956) and Massera (1956a) showing that asymptotic stability implies the existence of a Liapunov function and the theorems of Krasovskii (1963) and LaSalle (1960) refining Liapunov's sufficient conditions for asymptotic stability. The largely expository paper of Kalman and Bertram (1960) was important in showing how Liapunov theory could be used to help understand control problems and by the mid 1960s Liapunov methods were being used in a broad range of applications. (More can be found in the survey (Brockett, 1966).)

[11] The process is similar to the one later used in back-stepping (Kokotovic, 1992). In 1995 Fliess, Lévine, Martin, and Rouchon (1995) introduced a definition of flatness as a property of systems of the form $\dot{x} = f(x, u)$ characterizing a class of systems slightly larger than the feedback linearizable systems discussed here.

8.1. Global results

Although it may seem universally desirable to design systems that are asymptotically stable in the large, this goal may be impossible to achieve. Some background material is needed to explain this. Let M and N be oriented manifolds of the same dimension and without boundary.[12] If $f : M \to N$ is a mapping sending N onto M then $x \in M$ is said to be a *regular point* of f if the Jacobian of f is nonsingular at x. In contrast, $y \in N$ is said to be a *regular value* of f if each inverse image of y is a regular point. The *degree* of the map f is defined as the number of inverse images at which $\det(\partial f/\partial x)$ is positive minus the number of inverse images where $\det(\partial f/\partial x)$ is negative. It can be shown that this number does not depend on the choice of the regular value y and hence is a property of f (Milnor, 1956). The degree is a global idea but it can be used to define a property of an isolated equilibrium point of $\dot{x} = f(x)$ by observing that if $f(0) = 0$ then the mapping $y = f(x)/\|f(x)\|$ sends a small sphere centered at an equilibrium point to the unit sphere of the same dimension. The *index* of an equilibrium point is defined as the degree of this map. It is an easy calculation to show that the index of an asymptotically stable critical point in an n-dimensional setting is $(-1)^n$ and, more generally, if an equilibrium point is a regular point, the index is sign of the determinant of the Jacobian evaluated at that point. This theory is developed in considerable detail the book of Krasnoselskii and Zabreko (1984).

Let X be a compact, orientable manifold without boundary. Let f and g be vector fields defined on X such that the equilibria of $\dot{x} = f(x)$ and $\dot{x} = g(x)$ are isolated, The Poincaré–Hopf theorem (Milnor, 1956) says that the sum of the indices over all the equilibrium points for the two differential equations is the same.[13]

Example. Consider the differential equation $\dot{x}_1 = x_1x_2$; $\dot{x}_2 = -x_1^2$ viewed as evolving on the one-dimensional manifold defined by $\{(x_1, x_2)| = x_1^2 + x_2^2 = 1\}$. There are two equilibria, $(x_1, x_2) = (0, \pm 1)$. The solution $(x_1, x_2) = (0, -1)$ is asymptotically stable whereas the solution $(x_1, x_2) = (0, 1)$ is not. Because in one dimension stable asymptotically equilibria have index -1 and unstable equilibria have index 1, we see that the sum of the indices of this vector field, and hence any vector field with isolated equilibria on the circle, is zero. More generally, consider the $n-1$-dimensional sphere $X = \{x|x^Tx = 1; x \in \mathbb{R}^n\}$. If h is a unit vector, the differential equation $\dot{x} = -h + (x^Th)x$ evolves on X and has just two equilibria, $x = \pm h$. The index at the equilibrium point $x = h$ is $-(-1)^n$ whereas the index at the equilibrium point $x = -h$ is $(-1)^{n-1}$. Thus the sum of the indices of the equilibrium points of any vector field on S^{n-1} with isolated equilibria is 0 if n is even and 2 if n is odd. On the other hand, if $\Omega = -\Omega^T$ is nonzero, then the differential equation $\dot{\Theta} = \Omega\Theta$ on the orthogonal group has no equilibria so the sum of the indices for any smooth differential equation on the orthogonal group must be zero as well.

The Poincaré–Hopf theorem links local properties of the set of equilibria to the global structure of the manifold. In 1967 Wilson (1967) proved something more along these lines. He showed that the domain of attraction of an asymptotically stable equilibrium point is diffeomorphic to $\mathbb{R}^n$. The general idea is suggested by a simple example. Consider the manifold defined by $\{(x, y, z)|x^2 + y^2 + z^2 = 1\}$ and the control system defined by $\dot{x}_1 = x_2, \dot{x}_2 = -x_1 + ux_3; \dot{z} = -x_2u$. The control law $u = -x_1$ makes the point $(1, 0, 0)$ asymptotically stable in the sense that initial conditions close to this point will flow to it. In fact the solution from any initial condition except the initial condition $(-1, 0, 0)$ flows to $(1, 0, 0)$. However it is not possible to adjust this "near perfect" situation so as to have all initial values go to $(1, 0, 0)$. Because solutions of differential equations depend continuously on the initial conditions and there is no continuous map from the sphere to the neighborhood of a point global stability is impossible. This is an example of a global obstruction to stability. It took some time for such results to be absorbed and in the meantime a number of wrong claims about global stability of attitude control systems appeared. For example, see Mortensen (1968). Wilson also showed that a necessary and sufficient condition for M to be an invariant set which is asymptotically stable in the large for $\dot{x} = f(x)$, $x \in \mathbb{R}^n$ is that $\mathbb{R}^n - M$ be diffeomorphic to $S^{n-1} \times \mathbb{R}$.

8.2. Local stability and stabilization

Problems involving the design of a feedback control law that stabilizes an open loop unstable solution are among the most important problems in control. They stand in contrast to the problems motivating Poincaré and Liapunov, who wrote about the stability of immutable systems such as ships and heavenly bodies. Liapunov demonstrated in his thesis that exponentially stable linear systems have quadratic Liapunov functions and that if the system is time invariant the Liapunov function can be taken to be time invariant. Easy arguments based on solving $QA + A^TQ = -I$ for Q show, for example, that if f is twice differentiable and $f(0) = 0$, then for

$$\dot{x} = f(x) = Ax + f_1(x); \quad A = \left.\frac{\partial f}{\partial x}\right|_0$$

the stability properties of $\dot{x} = Ax$ determine the (local) stability properties of the nonlinear system, provided that A does not have eigenvalues with real parts zero.

It is often of interest to investigate the asymptotic stability of an equilibrium point of a differential equation $\dot{x} = f(x)$ written in $\mathbb{R}^n$ but defining a flow on a manifold described by $X = \{x|\phi_i(x) = 0, i = 1, 2, \ldots, k\}$. This is the case, for example, when there are one or more constants of motion such as energy, angular momentum, etc. Now the relevant linearization must take the constraints into account. Suppose that $f(x_0) = 0$ with $x_0 \in X$. Introduce matrices N and A, N being $n \times k$,

$$N = \left[\left.\frac{\partial \phi_1}{\partial x}\right|_{x_0}; \left.\frac{\partial \phi_2}{\partial x}\right|_{x_0}; \cdots \left.\frac{\partial \phi_k}{\partial x}\right|_{x_0} \right]; \qquad A = \left.\frac{\partial f}{\partial x}\right|_{x_0}.$$

Because $\phi(x)$ will only be constant if $\langle(\partial\phi_i)\partial x, f(x)\rangle = 0$, it follows that the null space of N^T is an invariant subspace for A. If the eigenvalues of A, restricted to this space have negative real parts, then x_0 is asymptotically stable.

There is an analogous result on "linear controllability" in the neighborhood of an equilibrium point obtained by similar reasoning. If $\dot{x} = f(x) + \sum u_ig_i(x)$ with $f(x_0) = 0$ defines a control system on $X = \{x|\phi_i(x) = 0, i = 1, 2, \ldots, k\}$ let $B = [g_1(x_0), g_2(x_0), \ldots, g_m(x_0)]$. Because x is assumed to evolve on X for all u, it follows that the system is locally controllable in the neighborhood of x_0 provided that $B, AB, \ldots, A^{n-1}B$ spans the null space of N.

As is well known, there exists a stabilizing linear, time invariant, feedback control law for any controllable linear time invariant system. This can be shown in a number of ways but perhaps the first full proof comes as a consequence of Kalman's 1960 paper on least-squares optimal control (Kalman, 1960b). Moreover, given a nonlinear system of the form $\dot{x} = f(x, u)$ with $f(0, 0) = 0$ and linearization $\dot{x} = Ax + Bu$, there exists a linear time invariant control law that makes the null solution (locally) stable provided that the linearized system is controllable. Brockett (1973b)

[12] A differentiable manifold is orientable if it can be covered with overlapping patches such that the changes of coordinates between the patches only involve maps with positive Jacobians. The Klein bottle is the standard example of a non orientable surface.

[13] The sum of the indices is the Euler Characteristic of X. Sometimes the easiest way to compute the Euler characteristic is to find a suitable vector field on X and find the sum of the indices. It is possible to substitute an assumption about in coming trajectories for the assumption of compactness; see Milnor (1956).

considers systems of the form $\dot{x} = Ax + uBx$ evolving on the manifold $\{x|x^Tx = 1\}$. Among the results proved there is the fact that a point $x_0 \in X$ is rendered asymptotically stable by the control law $u = -x_0^T B^T$, provided that the linear controllability condition requiring that the set $Bx_0, ABx_0, \ldots, A^{n-1}Bx_0$ span the orthogonal complement of $(\partial\phi/\partial x) = 2x_0$ is satisfied.

Example. Consider two different systems of the form $\dot{x} = (A + uB)x$ each evolving on a manifold of the form $X = \{x|x^Tx = 1\}$. Specifically,

$$A_1 + uB_1 = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & u \\ 0 & -u & 0 \end{bmatrix};$$

$$A_2 + uB_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & u & 0 \\ 0 & -u & 0 & u \\ 0 & 0 & -u & 0 \end{bmatrix}.$$

Lie bracket computations show that both systems are controllable.

$$\left[\begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}\right] = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

For the first system, the function $v(x) = x_3$ takes on a minimum at $x = [0, 0, -1]^T$ and its derivative is $-ux_2$. Thus the control law $u = x_2$ results in $\dot{v} = -x_2^2$, and renders the equilibrium point $x = [0, 0, -1]^T$ asymptotically stable. (Note that x_2^2 does not vanish identically along any solution except $x = [0, 0, \pm 1]^T$.) The domain of attraction is all of X except the point $[0, 0, 1]^T$. On the other hand, for the second system, the function $v(x) = x_4$ takes on a minimum at $x = [0, 0, 0, -1]^T$ and the control law $u = -x_3$ gives $\dot{v} = -x_3^2$, this choice does not result in asymptotic stability of the solution $x = [0, 0, 0, -1]^T$ because $\dot{v}$ vanishes identically along a nonzero solution. In fact, there is no control depending continuously on x that stabilizes the solution $x(t) \equiv [0, 0, 0, -1]^T$. Examples like these are explored in Brockett (1973b) where the derivative of a potential Liapunov function is evaluated as a function of the control and then the control is chosen to make it negative.

8.3. Critical cases in feedback stabilization

If $x(t) \equiv 0$ is a solution of $\dot{x} = f(x)$ and if $(\partial f/\partial x)|_0$ has one or more eigenvalues with zero real parts, then for the purposes of stability, this is called a *critical case*. In this situation the second and higher order terms are decisive in determining stability. Lossless physical systems, such as freely tumbling satellites with rotors and networks of coupled oscillators often fall into this case. Malkin (1959) devotes considerable space to this topic but no comprehensive theory exists. There is a corresponding idea for nonlinear control systems of the form $\dot{x} = f(x, u)$. If $f(0, 0) = 0$ and if the linearization about $x \equiv 0$, $u \equiv 0$ is $\dot{x} = Ax + Bu$, we say the stabilization problem corresponds to a *critical case* if A has eigenvalues with zero real parts that cannot be altered with linear feedback.

The existence of an open loop control that steers an arbitrary initial condition to zero is clearly a necessary condition for the existence of a stabilizing control law. Because controllability was known to be sufficient for the existence of a stabilizing control in the linear case, it was natural to ask if this condition is sufficient in more general situations or if something additional is needed. In Brockett (1983) it was shown that controllability is, in general, insufficient. Using the fact that an autonomous system with an asymptotically stable constant solution necessarily has a Liapunov function Kurzweil (1956), Massera (1956a), together with a suitable fixed point theorem, it was shown that a necessary condition for $\dot{x} = f(x)$ to have $x \equiv x_0$ as an asymptotically stable equilibrium solution is that f should map a neighborhood of x_0 onto a neighborhood of 0. This condition can also be interpreted in terms of the index defined above; if f is not onto then clearly the index is zero and not $(-1)^n$ as it would need to be for asymptotic stability.

Example. Consider the three-dimensional system

$$\dot{x}_1 = u_1; \qquad \dot{x}_2 = u_2; \qquad \dot{x}_3 = u_1x_2 - u_2x_1.$$

Expressing this as $\dot{x} = g_1(x)u_1 + g_2(x)u_2$, an easy calculation shows that $g_1, g_2, [g_1, g_2]$ span $\mathbb{R}^3$ and so this driftless system has the property that any state can be steered to any other state and it can be done in arbitrarily small positive time. However, upon linearizing about $(x, u) = (0, 0)$ we get an uncontrollable system so that the linearized system does not provide a basis for finding a stabilizing feedback law. Moreover, because the equations $[a, b, c] = [u, v, x_1v - x_2u]$ cannot be solved for all $[a, b, c]$ in a neighborhood of zero, the theorem referred to above shows that no continuous feedback control law stabilizes the origin. In fact, applying the theorem to any system of the form $\dot{x} = G(x)u$; $u \in \mathbb{R}^m$ with the dimension of u less than the dimension of x shows that continuous stabilization is impossible.

9. Carleman linearization and Volterra series

In 1932 Carleman (1932) described a linearization (or embedding) technique which, subject to some limitations, facilitates the construction of higher order approximations to analytic maps. When applied to the problem of approximating solutions of nonlinear differential equations, it provides a systematic way of extending the accuracy of approximations beyond that given by ordinary linearization. When used as an approximation technique, Carelman's procedure associates to an equation evolving in $\mathbb{R}^n$ an approximation evolving on an n-dimensional sub manifold of some higher dimensional cartesian space. In this sense it defines an embedding. Together with the results on Volterra series for bilinear systems cited above, it provides a constructive approach to the computation of Volterra series for a broad class of nonlinear systems. More generally, it sheds light on the Poincaré Liapunov linearization theorem and, in certain circumstances, gives a constructive approach to the theorem of Ado cited above.

9.1. Carleman linearization

To put the ideas in their simplest setting, assume that x is a scalar satisfying the differential equation $\dot{x} = f(x) + ug(x)$ with $x(0) = 0$ and $f(0) = 0$. Express the Taylor series expansion for $f + ug$ as

$$f + ug = ub_0 + (a_1 + ub_1)x + (a_2 + ub_2)x^2 + \cdots.$$

Letting $c_i = a_i + ub_i$ we then get a bilinear equation

$$\frac{d}{dt}\begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ c_0 & c_1 & c_2 & c_3 & \cdots \\ 0 & 2c_0 & 2c_1 & 2c_2 & \cdots \\ 0 & 0 & 3c_0 & 3c_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}\begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \\ \vdots \end{bmatrix}.$$

Truncating the series by setting terms of order $p + 1$ to zero, gives a bilinear system. Solving this using the Peano–Baker series, as discussed above, gives a correspondingly accurate approximation to the true solution. If $\dot{x} = f(x) + \sum u_i g_i(x)$ with x a vector, the form remains the same. If x has components $x_1, x_2, \ldots, x_n$ it is convenient to denote the vector consisting of all monomials of degree p in $x_1, x_2, \ldots, x_n$ as $x^{[p]}$. This is a vector of dimension $(n+p-1)!/(p-1)!(n-p+1)!$. If $f(0) = 0$ the Taylor series expansion can be written as

$$f + \sum u_i g_i = \sum u_i B_{0i} + \left(A_1 + \sum u_i B_{1i}\right) x^{[1]} + \left(A_2 + \sum u_i B_{2i}\right) x^{[2]} + \cdots.$$

In terms of this notation the Carleman linearization can be organized, as in the scalar case, as a differential equation

$$\frac{d}{dt}\begin{bmatrix} 1 \\ x^{[1]} \\ x^{[2]} \\ x^{[3]} \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ C_0 & C_{11} & C_{12} & C_{13} & \cdots \\ 0 & C_{20} & C_{21} & C_{22} & \cdots \\ 0 & 0 & C_{30} & C_{31} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} 1 \\ x^{[1]} \\ x^{[2]} \\ x^{[3]} \\ \vdots \end{bmatrix}.$$

Setting all powers of x above a certain level equal to zero can only give a meaningful approximation if the components of x are less than one in magnitude. This will be the case if $x(0) = f(0) = 0$ and u has a sufficiently small $L_1[0, T]$ norm. If the null solution of $\dot{x} = f(x)$ is exponentially stable this is even true for u small in the $L_1[0, \infty)$ norm.

The above approach was developed in detail in Brockett (1976b) together with results on the realization of Volterra series as discussed below. Subsequent papers by Agrachev and Gamkrelidze (1978), Crouch (1977), Fliess and Lamnabhi-Lagarrigue (1982), Fliess and Normand-Cyrot (1982), Lesiak and Krener (1978), and others examined these ideas, in some cases using theories based on noncommutative power series following Chen (1957). It is intuitive that a system having a convergent Volterra series can be approximated by a finite Volterra series. This idea is implicit in Krener (1975) and is discussed more explicitly in Brockett (1976b). Crouch (1977) undertakes the study of the natural state space associated with finite Volterra series and shows that it is always diffeomorphic to $\mathbb{R}^n$ for some n.

In this development there is no requirement that the system be time invariant; A and B_i can be time varying. However, if they are constant and if $f(x(0), 0) = 0$, then the kernels are shift invariant. Furthermore, if $\hat{x}(\cdot)$ is any nonzero solution of $\dot{x} = f(x)$; $x(0) = \hat{x}(0)$ then making the change $x \mapsto x - \hat{x}$ puts the analysis back into the above case.

9.2. Realization theory

For linear systems there is a very tight connection between the differential equation description and the description in input–output form via an integral representation, especially in the case of controllable and observable time invariant systems. In the case of linear analytic systems there are similar results relating the differential equation description to the Volterra kernels. It follows from the above construction that if f and g_i are time invariant, $f(0) = 0$ and $x(0) = 0$ then the kernels take the form

$$w_p = e^{A(t-\sigma_1)} B e^{A(\sigma_1 - \sigma_2)} B \cdots e^{A(\sigma_{p-1} - \sigma_p)} \xi$$

and that given a pth-order kernel of this form there exists a bilinear system $\dot{x} = Ax + uBx$ with this as its p^{th} kernel and with all other kernels zero. Thus the existence of realizations comes down to the existence of realizations of single kernels which can then be added together. In this sense, as already remarked above in connection with the work of Krener (1975), bilinear systems have strong approximation properties.

Consider the contrast between an "upper triangular" system as given by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_{n-1} \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1\,n-1} & a_{1n} \\ 0 & a_{22} & \ldots & a_{2\,n-1} & a_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & a_{n-1\,n-1} & a_{2n} \\ 0 & 0 & \ldots & 0 & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix}$$

and a "strictly upper triangular system" obtained when all the diagonal terms in the matrix are zero. In the latter case, the solution is expressible in terms of nested integrals,

$$x_n(t) = x_n(0); \qquad x_{n-1}(t) = x_{n-1}(0) + \int_0^t a_{n-1n}(\sigma)\, d\sigma\; x_{n-3};$$
$$x_{n-3}(0) + \cdots$$

whereas in the more general situation the solution for x_n involves $\exp \int_0^t a_{nn}\sigma\; d\sigma$ and the expressions for the remaining components of x, while expressible explicitly, are correspondingly more complex. However, the change of variables

$$z_i = \exp\left(-\int_0^t a_{ii}(\sigma)\, d\sigma\right) x_i$$

converts the upper triangular system in x to a strictly upper triangular system in z.

Remark. Observe that for $\dot{x} = uy + uz$; $\dot{y} = uz$; $\dot{z} = 0$ we have $y = \int u dt$; $x = \int u \int u d\sigma_1 d\sigma_2 = (1/2)\left(\int u d\sigma\right)^2$. However, it is not possible to realize the input–output relation $y(t) = \int_0^t u^2(\sigma) d\sigma$ by means of a system of the form $\dot{x} = f(x) + ug(x)$.

9.3. Explicit short time approximations

Consider a time varying nonlinear system $\dot{x} = f(x, t) + u_i g_i(x, t)$; $x(0) = x_0$ and suppose that the solution for $u = 0$ is $\hat{x}$. It is assumed that the system $\dot{x} = f(x)$ with the given initial condition has no finite escape time on $[0, t_1)$ where t_1 may be infinity. There are two types of higher order approximation to the input–output behavior that have found use. One is based on expanding f and g_i in a Taylor series about $x = \hat{x}$ and this leads to a time varying bilinear system as described above. The time interval over which the Volterra series represents the solution is limited by t_1 and by the size of the L_1 norm of u. A second type of high order approximation proceeds by introducing an additional equation $\dot{x}_0 = 1$; $x_0(0) = 0$. It then works with an extended system characterized by $(dx_e)/dt = f_e(x_e) + u_i g(x_e)$, where f_e and g_e and their derivatives with respect to x are obtained by replacing the time dependence of f and g by a dependence on x_0 and then expanding them in a Taylor series in x and x_0. This approximation is, of course, only valid for small u and small $|t|$.

To be explicit, consider the system in $\mathbb{R}^n$ evolving as $\dot{x} = f(x) + \sum u_i g_i(x)$. Suppose that the Taylor series expansion for the solution with $u \equiv 0$ is given by $\hat{x}(t) = a_0 + a_1 t + a_2 t^2 + \cdots$. Introduce $z = x - \hat{x}$ and $z_e = [t, x - \hat{x}^T]^T$. Thus we can write

$$\dot{z}_e = \begin{bmatrix} 1 \\ f(z - \hat{x}) \end{bmatrix} + \sum u_i \begin{bmatrix} 0 \\ g_i(z - \hat{x}) \end{bmatrix}.$$

Although $\hat{x}$ depends on t we can replace t by the first component of z_e and in that way expand $f(z - \hat{x})$ in a power series in z_e. The result is a differential equation of the form $\dot{z}_e = \hat{f}(z_e) + \sum u_i \hat{g}_i(z_e)$ now with $z_e(0) = 0$. In this way one can get a Volterra expansion for the original system which is accurate to arbitrary order but because $z_0 = t$ this is a small time approximation.

10. Variational theory and optimal control

As discussed above, subsets of $\mathbb{R}^n$ described as $X = \{x|\phi_1(x) = 0, \phi_2(x) = 0, \dots, \phi_k(x) = 0\}$, can be given the structure of a m-dimensional manifold if the matrix

$$M = \begin{bmatrix} \partial\phi_1/\partial x_1 & \partial\phi_1/\partial x_2 & \cdots & \partial\phi_1/\partial x_n \\ \partial\phi_2/\partial x_1 & \partial\phi_2/\partial x_2 & \cdots & \partial\phi_2/\partial x_n \\ \cdots & \cdots & \cdots & \cdots \\ \partial\phi_k/\partial x_1 & \partial\phi_k/\partial x_2 & \cdots & \partial\phi_k/\partial x_n \end{bmatrix}$$

is of rank $n-m$ for all x in some open set of $\mathbb{R}^n$ containing X. If ψ is a differentiable function $\psi : \mathbb{R}^n \to \mathbb{R}^1$, it is also differentiable when considered to be a map from X to $\mathbb{R}^1$. The condition for $x_0 \in X$ to be a stationary point for ψ, as a map $\psi : X \to \mathbb{R}$ is that there exist real numbers λ_i such that

$$\frac{\partial}{\partial x}\left(\psi(x) + \sum \lambda_i\phi_i(x)\right) = 0.$$

That is, if X admits the structure of an m-dimensional manifold then the first order necessary condition for the mapping $\psi : X \to \mathbb{R}$ to be stationary at x_0 can be expressed in terms of Lagrange multipliers.

Subject to similar limitations, it is possible to develop the Euler–Lagrange equations for constrained systems. The argument starts with the usual Taylor series expansion followed by integration-by-parts argument applied to the first order approximation,

$$\int_0^T L(\dot{x} + \dot{\delta}, x + \delta,)\; dt$$
$$\approx \int_0^T L(\dot{x}, x)\; dt + \int_0^T \left\langle \delta, -\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} + \frac{\partial L}{\partial x}\right\rangle dt + \left\langle \delta(t), \left.\frac{\partial L}{\partial \dot{x}}\right|_0^T \right\rangle.$$

This then leads to the first order necessary conditions

$$\left\langle \delta, \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x}\right\rangle = 0 \quad \text{if} \left\langle \delta, \frac{\partial \phi_i}{\partial x}\right\rangle = 0.$$

Because of the constraints it is necessary to limit the variations. Subject to conditions analogous to the rank condition imposed on M above, this can be expressed as saying there exist (generally time varying) multipliers λ_i such that the second order equation

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} - \sum \lambda_i \frac{\partial \phi_i}{\partial x} = 0$$

characterizes the possible optimal trajectories. This goes back to Lagrange himself and has found wide use in mechanics where λ has the interpretation of a force, if applied to an unconstrained particle, would result in a motion that satisfies the constraints.

The Euler–Lagrange approach to variational problems leads to a system of second order differential equations in x, the "configuration variables". This is in contrast with the maximum principle which leads to a system of first order equations involving configuration variables and linear functionals on $\dot{x}$. The differential geometric distinction is that the Euler–Lagrange approach expresses the first order necessary conditions in terms of a vector field on the tangent bundle whereas the maximum principle formulates the first order necessary conditions in terms of a vector field on the cotangent bundle.[14]

[14] The *tangent bundle* associated with a manifold X can be thought as the set of all pairs $(x, \dot{x})$ with $x \in X$ and $\dot{x}$ a vector tangent to X at the point x. The *cotangent bundle* is the set of pairs (x, p) with $x \in X$ and p being a linear functional assigning a real number to each $\dot{x}$ tangent to X at x. If X is n-dimensional then these spaces are $2n$-dimensional.

The maximum principle can be applied to a system evolving on an embedded manifold $\{x|\phi_i(x) = 0\}$ in the usual way, i.e., when the problem is formulated so that there is no running cost and $\dot{x} = f(x, u)$ then $h = p^T f(x, u)$ etc., letting the manifold constraints remain implicit. However, because $\phi_i(x) = 0$ for all choices of u, adding a function $\mu(\phi_1(x), \phi_2(x), \dots, \phi_k(x))$ to the Hamiltonian does not change the optimal trajectories. Even though replacing h by $h + \mu(\phi_1(x), \phi_2(x), \dots, \phi_k(x))$ leaves the x-trajectories unchanged, the presence of μ is reflected in the co-state equation. Specifically, the co-state equation is now

$$\dot{p} = -\frac{\partial}{\partial x}h(x, p) - \frac{\partial}{\partial x}\phi_i(x).$$

Thus, when there are constraints p is not unique. A natural way to identify a unique p is to ask that it be perpendicular to the vectors $\frac{\partial \phi_i}{\partial x}$. This has the effect of making (x, p) an element of the cotangent space.

Example. Let X be an orthogonal matrix evolving as $\dot{X} = \Omega(u)X$ with $\Omega(u)$ skew-symmetric; in this case the Hamiltonian is $\text{tr}(P^T\Omega X) = \text{tr}(XP^T\Omega)$. The co-state equation, ignoring the implicit relationship $X^TX = I$, is $\dot{P} = -\Omega^TP$. However, because Ω is skew-symmetric, the Hamiltonian only depends on the skew-symmetric part of XP^T. Thus the solution of the equation for the skew-symmetric matrix $XP^T - PX^T$,

$$\frac{d}{dt}\left(XP^T - PX^T\right) = [\Omega, (XP^T - PX^T)]$$

captures the relevant information in P and $(X, (XP^T - PX^T))X = (X, XP^TX - P)$ defines a point in the cotangent bundle.

An early formulation of optimization of trajectories on manifolds was given by Albrecht (1968). Closer to the problems we discuss here, Griffiths (1983) gives a purely differential geometric formulation of variational methods based on the exterior calculus and cites some work in control as being a good source of problems. Brockett (1973b) applies the maximum principle to a class of problems for which the state evolves on Lie groups and homogeneous spaces.

One further general point that should not to be overlooked. The maximum principle can be viewed as showing that a system loses controllability in the neighborhood of an optimal trajectory. For example, consider a fixed time problem on $[0, t_1]$ with a performance measure, expressed in terms of the end point, $\phi(x(t_1))$. If $\frac{\partial \phi}{\partial x}$ is nonzero and x^* is an optimal trajectory then a variation δ in $x(t_1)$ results in a first order change in ϕ given by $\langle(\partial\phi/\partial x), \delta\rangle$. If x^* minimizes ϕ then δ cannot be arbitrary so the system is not controllable around x^*.

10.1. Shortest path problems

Some of the most widely studied problems in the calculus of variations are associated with finding the shortest distance between two points in a Riemannian manifold. Given a manifold X and a positive definite quadratic form G, expressed in terms of a local coordinate system as $g_{ij}(x)$, measure distance by

$$d(x_1, x_2) = \min \int_0^1 \sqrt{\sum dx_i g_{ij} dx_j}; \qquad x(0) = x_1; \qquad x(1) = x_2.$$

In $\mathbb{R}^n$ with $G = I$ the shortest paths are straight lines, on spheres they are great circles, etc. The study of geodesics is facilitated by first showing that the above formulation using the square root in the integrand leads to the same paths as the analytically simpler problem of minimizing

$$\eta = \min \int_0^1 \sum \dot{x}_i g_{ij}(x)\dot{x}_j\; dt; \quad x(0) = x_1; \quad x(1) = x_2$$

with the minimum taken over some class of differentiable paths. Factoring G^{-1} as $G^{-1} = BB^T$, the geodesic problem can be phrased as the following least squares control problem. Given $\dot{x} = B(x)u$, find a u that steers $x(0) = x_1$ to $x(1) = x_2$ while minimizing $\int_0^1 u^T u \, dt$.

Applying the maximum principle, with a Hamiltonian $\sum p^T b_i u_i + (1/2)u_i^2$ we see that the optimal choice of u is $u = -B^T p$ and the resulting state–costate equations are

$$\dot{x} = BB^T p; \qquad \dot{p} = -\frac{1}{2}\frac{\partial}{\partial x} p^T B^T B p.$$

This development assumes that the problem is normal, which it is because $G(x) > 0$. Nonlinear control theory has nothing particularly new to say about the Riemannian geodesic problem. The novelty comes from the fact that the minimization problem makes sense more generally. Instead of asking that $B^T B = G^{-1}$ be positive definite, what is required is that the system $\dot{x} = B(x)u$ be controllable.

An interesting example, apparently first solved by Baillieul (1975) involves the orthogonal group. Let $\mathcal{X}$ be the space of 3-by-3 orthogonal matrices and consider the system described by

$$\dot{X} = \begin{bmatrix} 0 & u_1 & u_2 \\ -u_1 & 0 & 0 \\ -u_2 & 0 & 0 \end{bmatrix} X.$$

In this case the manifold X is three dimensional and the control space is two dimensional. If we wish to minimize the integral of $u^2 + v^2$ subject to $X(0) = X_0$ and $X(1) = X_1$ we have a typical sub Riemannian geodesic problem. In this case the optimal path takes the form $X(t) = e^{\Omega t} e^{(H-\Omega)t} X(0)$ with $\Omega = -\Omega^T$ and $H + H^T$. This is a prototype for a general class of geodesic problems on symmetric spaces. Among the more easily solved problems of this type is the first bracket controllable system investigated by Brockett (1981b) and Gaveau (1977). This expository paper of Strichartz (1986) surveys this material from a mathematical point of view, describing some relationships with other geometries and symmetric spaces.

10.2. Optimality of bang–bang controls

In the early 60s various bang-coast-bang controls were found to be optimal for linear systems with either a penalty or a constraint on fuel consumption, such as the famous Lawden spiral Lawden (1962), but because the time-optimal control of linear time invariant systems with the input vector constrained to a closed set had been shown to be bang–bang, it was natural to investigate conditions under which a similar result might hold in nonlinear settings. (Sussmann, 1972) established such a result for bilinear systems evolving on $Gl(n)$, but only if the matrices in the bilinear system commute. Krener (1971) gives sufficient conditions on a class of nonlinear control systems for an optimal control to be bang–bang using higher order control variations generalizing those developed by Kelley and others. A few years later Sussmann gave conditions for a linear analytic system with bounded controls to have optimal solutions that are bang–bang. His conditions involve comparisons between the size of the ad-chain terms, $ad_f^k(g)$, and the size of brackets involving brackets such as $[g_i, g_j]$.

10.3. Higher order optimality conditions

Of course no general explicit solutions are available for $\dot{x} = f(x) + \sum g_i(x)u_i$ but for some purposes an analysis of the linearized system is sufficient. Indeed, linearization about a presumed optimal trajectory provides the basis for the Euler–Lagrange equations and the maximum principle. However, a closer examination of various applied problems revealed cases where the first order approximations are indecisive and higher order approximations are needed in order to characterize optimal trajectories. Early work here includes the work covered by the Kelly–Kopf–Moyer conditions (Kelley et al., 1966). The book of Bell and David (1965) gives an idea about the state of affairs circa 1965 and Gabasov and Kirillova (1972) survey the field as of 1972.

Krener's thesis (Krener, 1971) and his paper (Krener, 1977) develop a more complete and systematic analysis of higher order optimization theory, based on differential geometric thinking and over the decade of the 70s such ideas were explored in great detail. As is the case when characterizing the extreme values of a smooth function defined on a subset of $\mathbb{R}^n$, one uses a Taylor series expansion and investigates the lowest order nonzero term. If it is of full rank and if there are no endpoint constraints things are straight forward. In the control setting, an expansion of the performance measure in a Volterra series provides a straight forward approach and the use of a second order Volterra expansion in this way is described in Brockett (1976b). When there are endpoint conditions on the state the situation is more complicated.

11. Stochastic processes on manifolds

The use of stochastic models in physics and engineering has a long history, being basic to the subjects of information theory, communication technology, engineering thermodynamics, and, reaching back further in time, statistical mechanics. Many of the most compelling results from the early 20th century, such as the measurement and explanation of the power spectrum of black body radiation and its relationship to the Rayleigh–Jeans law, involve stationary processes with non equilibrium processes playing no role. Beginning in the 1940s, inspired by the work of Kolmogorov and Weiner on time series, engineers interested in communications and control began to adapt work on stationary processes to a wider set of purposes. In the area of stochastic control this involved finding controllers that give optimal steady state performance for linear systems with Gauss–Markov statistics, with 'optimal' being interpreted in an expected value sense. The textbook by Newton (1957) is a well-known example of how these ideas were put to use in a control theoretic setting.

However, it soon became clear that many important applications involved nonlinear effects and non stationary processes. Initial efforts to treat such problems took a discrete time approach (e.g. the Kalman filter), building on classical tools in probability theory. The landmark paper of Itô (1946) deals with continuous time processes of the type often found in physics and engineering situations, providing a deeper understanding of the meaning and use of stochastic differential equations. This paved the way for many new developments in continuous time stochastic control. In particular, his work helped to clarify the relationship between the measure theoretic approaches developed in the mathematical literature and the stochastic differential equation formalism that had been used in much of the physics and engineering literature up until that point.[15] McKean's 1969 book (McKean, 1969) develops Itô's ideas in a more leisurely way, making the subject accessible to a wider audience.

15 We refer to the physics literature as representing the Stratonovich point of view because in his book Stratonovich (1963) he provided a detailed description of the central difference interpretation of stochastic integration, although many examples of its use predate his work. Insofar as calculations are concerned, the relationship between these two formalisms come down to the choice between writing $\eth x = f(x)dt + g(x)\eth w$ and using the Stratonovich calculus and corresponding expectation rule vs. describing the system as $dx = f(x)dt + (1/2)\partial g/\partial x) g dt + g(x)dw$ and using the Itô calculus and the Itô expectation rule.

11.1. Nonlinear stochastic differential equations

Let w be a standard Wiener process. In his paper Clark (1973) summarized some of the basic results needed to explain the meaning of stochastic differential equations in Itô form

$$dx = f(x)dt + \sum_{i=1}^{m} g_i(x)dw_i$$

when x is to be thought of as evolving on a manifold $X \subset \mathbb{R}^n$ defined as $X = \{x|\phi_i(x) = 0, i = 1, 2, \ldots, k\}$. As a consequence of the Itô differentiation rule

$$d\phi_i = \left\langle \frac{\partial \phi_i}{\partial x}, f(x)dt + \sum_{i=1}^{m} g_j(x)dw_j \right\rangle + \frac{1}{2}\sum_{i=1}^{m} g_i^T \frac{\partial^2 \phi_i}{\partial x_j x_k} g_k dt$$

it is necessary that for $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, m$, both $\langle \partial\phi_i/\partial x, g_i\rangle = 0$ and

$$\left\langle \frac{\partial \phi_i}{\partial x}, f \right\rangle + \left\langle \frac{\partial \phi_i}{\partial x}, \frac{1}{2}\sum_{k=1}^{m} \frac{\partial g_k}{\partial x} g_k \right\rangle = 0.$$

A widely studied example of a stochastic equation evolving on an embedded manifold is the conditional density equation derived by Wonham (1965). In this case the system evolves on interior of the $n-1$-dimensional standard simplex in $\mathbb{R}^n$. The setting for this problem is this. Let x be a finite state jump process taking on values in the set of real numbers labeled $\{d_1, d_2, \ldots, d_n\}$. Suppose that in the absence of any observations, the probability that x takes on the value x_i is p_i and $\dot{p} = Ap$ describes the evolution of p. Given this, one brings observation into the picture. Assume that x is observed with additive white noise, i.e., one makes available $dy = xdt + d\nu$ with ν being a Wiener process. The evolution of p, conditioned on the past of y, then satisfies the Itô equation

$$dp = Apdt + (D - e^T DpI)(dy - (e^T Dp)dt)$$

where e is an n-vector of all ones and D is a diagonal matrix with diagonal elements d_i.

Some examples from physics and engineering involving stochastic processes evolving on manifolds such as spheres, orthogonal groups, etc., are reviewed and referenced in Brockett (1973a).

11.2. Multiplicative noise models

Some of the most widely used stochastic equations arising in applications involve linear systems with stochastic coefficients—a stochastic analog of bilinear systems. If the coefficients are general Gauss–Markov processes analysis is difficult but in the "white noise" coefficient case, i.e., when the equation takes the form

$$dx = Axdt + \sum dw_i B_i x$$

analytic methods have been very successful. Especially important examples emerging in the 1970s include the widely used option pricing model of Black and Scholes (1973); Merton (1973) and the Kossakowski–Lindblad models (Kossakowski, 1972; Lindblad, 1976) describing the loss of coherence in quantum mechanical systems. Two appealing aspect of multiplicative white noise model are: (i) the evolution equation for the moments of any order are linear and, (ii) such equations, like the bilinear equations discussed above, can describe systems evolving on manifolds such as spheres and Lie groups.

If x satisfies the Itô equation $dx = Axdt + Bxdw$ then the expected value of x satisfies

$$\frac{d}{dt}\mathcal{E}x = A\mathcal{E}x$$

and the second moment of x satisfies the linear equation

$$\frac{d}{dt}\mathcal{E}xx^T = A\mathcal{E}xx^T + \mathcal{E}xx^T A^T + B\mathcal{E}xx^T B^T.$$

This last equation illustrates the potentially destabilizing effect coming from the noise; even if A has eigenvalues with negative real parts the $B\mathcal{E}xx^T B^T$ term can lead to a growth in the second moment. The equations for moments of arbitrary order are given in Brockett (1976c) using ideas from multilinear algebra.

A widely used model of this type comes from quantum mechanics. The Schrödinger equation is linear and, in its basic form, it evolves with constant norm. However, in the study of open quantum systems it is often replaced by a stochastic equation of the form, $dx = Axdt + \sum B_i x dw_i$, with the equation being interpreted in the Stratonovich sense. With this interpretation the second moment satisfies

$$\frac{d}{dt}xx^T = Axx^T + xx^T A^T + \frac{1}{2}\sum B_i^2 xx^T - B_i xx^T B_i^T + xx^T (B_i^2)^T.$$

Moreover, if A and B_i are skew-Hermitian, as they are in quantum mechanics, then this equation takes the form

$$\frac{d}{dt}\mathcal{E}xx^T = [A, \mathcal{E}xx^T] + \sum [B_i, [B_i, \mathcal{E}xx^T]].$$

The quantum mechanical density matrix ρ arises in this way with $\rho = \mathcal{E}xx^T$; it plays a basic role in studying decoherence and is associated with the names Kossakowski and Lindblad. Written as $\dot{\rho} = -[ih, \rho] + [B, [B, \rho]]$ it plays a central role in the analysis and optimal control of quantum systems.

11.3. Controllability and Hörmander's hypoellipticity criterion

Beginning with the work of Einstein and Smoluchowski in the early 1900s, the study of stochastic process has been linked to the study of partial differential equations describing diffusion. In the simplest cases the connection is quite direct. The stochastic differential equation $dx = dw$ goes hand in hand with the one-dimensional heat equation

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

describing the evolution of the corresponding probability density of x. More generally, the diffusion equation corresponding to an Itô equation of the form $dx = f(x)dt + \sum_{i=1}^{m} g_i dw_i$ takes the form $\partial\rho/\partial t = L\rho$ with

$$L\rho = -\left\langle \frac{\partial}{\partial x_i}, f\rho \right\rangle + \sum_{i=1}^{m} \operatorname{tr}\left(\nabla\nabla^T g_i g_i^T\right)\rho; \quad \nabla\nabla^T = \left(\frac{\partial^2}{\partial x_j \partial x_k}\right).$$

Because the number of noise terms is typically less than the dimension of x, the operator L is often *degenerate* in the sense that the rank of the quadratic form defined by $\sum g_i g_i^T$ is less than the dimension of x. However, as was pointed out early on by Elliott (1971), a theorem of Hormander (1967) precisely characterizes the support of the solution of degenerate equations of the form $\partial\rho/\partial t = L\rho$ in terms of a Lie algebraic condition. This theorem provides a direct link between controllability ideas and the support of the solution of the Fokker–Planck equation. The relationship between controllability and positive definiteness of the variance in the case of linear systems is, of course, well known. A set of nonlinear equations illustrating these ideas,

$$\dot{x} = u; \qquad \dot{Z} = xu^T - ux^T; \qquad x \in \mathbb{R}^n; \qquad Z = -Z^T$$
$$dx = dw; \quad dZ = xdw^T = dwx^T$$

has been investigated in considerable detail by Brockett (1981b) and Gaveau (1977) and Brockett (1984).

11.4. Lie algebras and the conditional density

Given a stochastic process generated by an Itô equation of the above form, and given an observation equation, $dy = h(x)dt + d\nu$ with ν a standard Wiener process, the fundamental question in causal estimation theory is that of determining the conditional probability of $x(t)$ given the observation of y up until time t. Of course in the case of linear gaussian models this question has a relatively simple answer because the conditional density is gaussian and therefore characterized completely by its mean and variance. In more general settings the characterization of the conditional density, in terms of a differential equation for its evolution, is mathematically delicate. Given the success of the Kalman–Bucy filter, it is clearly of interest to know if there are nonlinear situations such that the conditional density can be propagated by means of a finite set of stochastic differential equations driven by the observations. That is, are there classes of systems beyond the linear gaussian model, which lead to finite dimensional filters and if so, how can they be characterized. In the spirit of the work of Wei and Norman (cited above), characterizing the intrinsic complexity of finding a solution to a time varying linear system $\dot{x} = \sum \phi_i(t)A_i x$ in terms of the Lie algebra generated by the $\{A_i\}$, the work of Brockett (1980) and Brockett and Clark (1980) applied this idea to a state estimation problems and this was pursued more generally in Brockett (1981a). Beneš (1981) gave a nice example involving systems with a hyperbolic tangent function as the drift term. Questions about the Lie algebras generated by the conditional density operators received a great deal of attention in the early 1980s and a number results establishing the intrinsic infinite dimensionality of many classes of problems were established.

12. Applications

Alongside the theory discussed above there were, in this period, a large number of real world problems being addressed. These were important in shaping the theory and in maintaining interest in the more theoretical developments. Among the earliest of these were problems involving the dynamics of satellites with internal dissipation and the design of attitude control systems appropriate in such circumstances. The growing field of robotics required the analysis of nonlinear models involving movements over a large work space. Problems involving the inverse dynamics are central and, in almost all situations, linearization is of limited use. Models for low loss energy conversion, both in portable electronic devices and as used in the electricity grid, are necessarily nonlinear. Work on nuclear magnetic resonance spectroscopy provided an extremely rich set of problems involving active sensing in a noisy environment. Although we do not go into such examples in great depth, we want provide at least a rough sketch of how the theoretical work described above is connected to applications. The 1974 survey paper of Bruni et al. (1974) discusses a wide range of concrete problems which are well characterized by bilinear equations, providing further examples.

12.1. Attitude control of satellites

Accurate attitude control is a basic requirement for the satellites used for communication and measurement of the earth's vegetation and natural resources. In addition, precise attitude control is essential for the success of scientific experiments such as the Hubble telescope and the probes for investigating the atmosphere of other planets in our solar system. These are excellent examples of the application of control engineering and especially its nonlinear aspects; in many cases the equations of motion are accurately known and on board computation is available to implement sophisticated algorithms. Actuation mechanisms include gas jets, reaction wheels and magnetic coils suitable for interacting with ambient magnetic fields. Here the need for a global theory is clear. This claim is well supported by the famous case of the Explorer 1 satellite (1958) which, on the basis of a linear analysis, was predicted to spin stably around an axis of minimum moment of inertia but, because of unmodeled nonlinear effects, ended up in a quite different state. More generally, a satellite may reach its desired orbit tumbling with an arbitrary orientation and an unpredicted angular momentum; the control system then needs to control its orientation such that as it moves around the earth a particular antenna remains pointed at the earth. Linearization about an equilibrium solution does not provide an adequate basis for a complete analysis.

In spin stabilization, as in "dual spin" satellites, and in attitude control more generally, the control mechanism is implemented using reaction wheels or other internal degrees of freedom. A large literature devoted to such questions emerged in the mid 1960s of which we mention work of Hooker and Margoulis (1965), Likins (1967), Meyer (1966) and Roberson and Wittenburg (1968). In various ways and with various problems in mind, these papers formulate attitude control and stabilization problems using reaction wheels and other means to provide torque. For example, they develop the equations of motion for a rigid body with controllable internal rotors in the form $\dot{x} = f(x) + \sum g_i(x)u_i$ with the u_i representing the torques applied to the reaction wheels (See Fig. 1). This type of work was subsequently extended by others including (Crouch & Bonnard, 1980; Krishnaprassad, 1985) in such a way to put more emphasis on controllability questions. One of the early practical uses of a system using an internal momentum wheel was the 1975 RCA dual spin satellite. See Hurbert (1981).

These equations involve three body angular velocities, three reaction wheel angular velocities and three parameters representing the orientation. However, by virtue of conservation of the total angular momentum, they are restricted to a six-dimensional sub manifold.

Putting the equation for Θ aside for a moment, the equations for the angular velocities can be written as

$$\dot{\omega}_1 = \frac{I_2 - I_3}{I_1}\omega_2\omega_3 + \frac{\omega_3 h_2 - \omega_2 h_3}{I_1} + \frac{u_1}{I_1}; \qquad \dot{h}_1 = u_1$$
$$\dot{\omega}_2 = \frac{I_3 - I_1}{I_2}\omega_1\omega_3 + \frac{\omega_1 h_3 - \omega_3 h_1}{I_2} + \frac{u_2}{I_2}; \qquad \dot{h}_2 = u_2$$
$$\dot{\omega}_1 = \frac{I_1 - I_2}{I_3}\omega_1\omega_2 + \frac{\omega_2 h_1 - \omega_1 h_3}{I_1} + \frac{u_3}{I_3}; \qquad \dot{h}_3 = u_3.$$

In an abbreviated notation, $\dot{\omega} = f(\omega, h) + I^{-1}u$; $\dot{h} = u$. The time derivative of $\frac{1}{2}\sum(I_i\omega_i^2 + h_i^2 J_i^{-1})$ is $\sum \omega_i u_i + h_i u_i J_i^{-1}$. In Brockett (1976a) the controllability of the three-dimensional Euler equations was investigated under various assumptions on the types of control available. An analysis of the controllability including the attitude equations was undertaken by Crouch and Bonnard (1980). A second type of control involving the possibility of controlling the system by changing the inertial tensor is also of interest. This was addressed by Kane and Scher (1969) in the context of the acrobatics of a falling cat and was subsequently investigated extensively using differential geometric ideas.

12.2. Nonholonomic mobile robotics

Interest in nonholonomic systems goes back to the nineteenth century, with early work by Boltzmann, Hamel and Appell among others. Whittaker's classic text (Whittaker, 1927) devotes considerable space to nonintegrable relationships in mechanics and contains references to these authors. However, it was only in the 1960s and 1970s, with the appearance of the book by Neimark and Fufaev

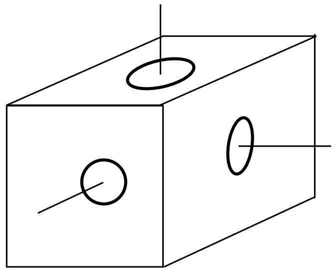

Fig. 1. Illustrating a rigid body with three reaction wheels aligned with the principal axes.

(1972), that there was a general appreciation of the wide scope of these ideas.[16]

In particular, the description of the kinematics of wheeled vehicles is the subject of many papers dealing with applications as varied as parking of automobiles and path planning for formations of mobile robots. One of the earliest papers of this type is the widely cited paper of Dubins (1957) dating from 1957. It considers the nonholonomic kinematics of an ordinary four wheeled automobile, focusing on shortest length paths, given an initial and final configuration. The unicycle is also a popular example. Its equations of motion can be expressed in terms of the coordinates of the center of the wheel, (x, y), and θ, the heading angle. They take the form of a linear analytic system,

$$\dot{x} = u_1 \sin\theta; \qquad \dot{y} = u_2 \cos\theta; \qquad \dot{\phi} = u_1.$$

A simpler example, illustrating the role of nonlinear controllability in robotic path planning, deals with a three wheeled tricycle-like vehicle operating on a level surface and powered by the front wheel. The geometry is illustrated in the figure on the left shown below. Let (x_1, x_2) denote the cartesian coordinates of the center of the front wheel and let θ denote the angle that the line segment joining the center of the front wheel to the midpoint back axle makes with the x axis, measured clockwise. The line segment is assumed to be of unit length. Let u_1 and u_2 denote the x and y components of the velocity of the center of the front wheel. The kinematic equations are

$$\dot{x}_1 = u_1; \qquad \dot{x}_2 = u_2; \qquad \dot{\theta} = u_2 \cos\theta + u_1 \sin\theta.$$

If the right-hand side is written as $g_1 u_1 + g_2 u_2$ then $[g_1, g_2] = -\partial/\partial\theta$. The Lie algebra generated by g_1 and g_2 is just three dimensional and is isomorphic to that of the two-dimensional euclidean group. The resulting distribution spans and, because there is no drift term, this system is controllable on $\mathbb{R}^2 \times S^1$. However, because there may be obstacles cluttering the space and limiting the motion, it is of interest to investigate possible limitations on the paths that can be followed. This raises the question of path following for nonholonomic systems.

The equations of motion for articulated vehicles, such as the tricycle with a trailer shown on the right-hand side of the figure, require additional state variables. In the case of the system depicted on the right-hand side of Fig. 2, the kinematic equations are four dimensional and the vector fields $g_1, g_2, [g_1, g_2]$ and $[g_1, [g_1, g_2]]$ span. It is of interest to generate a "shortest path" between two points in (x, y, θ)-space where shortest is defined as the integral of some function of x, y, θ, u_1, u_2. This is typical of the kind of path planning problems in which nonholonomic constraints lead to sub Riemannian problems as discussed above.

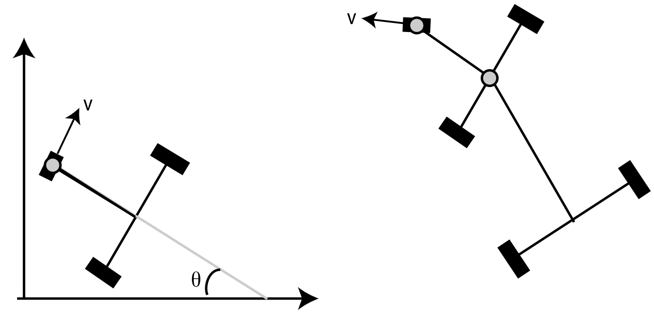

Fig. 2. Illustrating the kinematics of a tricycle (left) and a tricycle pulling a wagon (right).

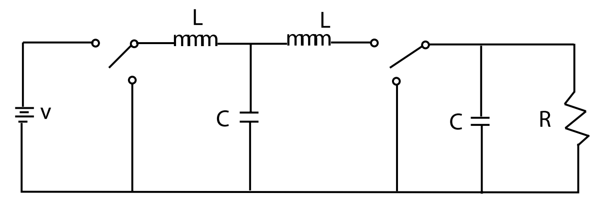

Fig. 3. Lossless voltage conversion.

12.3. Low loss electrical energy conversion

Circuits for lossless, or nearly lossless, conversion of electrical energy at one voltage to electrical energy at a second voltage find wide use in applications ranging from battery operated electronics to automobiles. They are also an essential component of most systems coupling renewable energy sources to the electric power grid. Of course in the case of alternating current this is routinely done with transformers but for conversion of direct current to direct current, a situation of growing importance, the process is more involved. The design and analysis techniques of Middlebrook and C'uk (1977) are frequently cited. See Sanders et al. (1991). The schematic shown (See Fig. 3) illustrates the power handling components of a typical circuit used to solve such problems. The essential parts are the inductors and capacitors, modeled as lossless passive components, the switches, a voltage source and a resistive load. The diagram does not show the circuitry needed to control the switches. Using inductor currents and capacitor voltages as state variables, standard circuit analysis techniques yield a differential equation model of the form

$$\dot{x} = Ax + \sum u_i B_i x + \sum u_i b_i$$

where each u_i represents the position of one particular switch and takes on just two values, which may be taken to be 1 and 0. In a typical situation the goal is to have the voltage of the right-most capacitor close to some desired level. The controllability properties of the system are clearly important in determining if this is possible.

There are two approximation schemes that are widely used to analyze such systems. The most elementary of these is simple averaging. Starting from the given equation and assuming that the switching policy is periodic in time, let $\bar{u}_i$ be the time average value of u_i. It is not to difficult to see that if the switching frequency is large compared with the natural frequencies of the system then the average value of ux is approximated by $\bar{u}\bar{x}$ and

$$\bar{x} \approx \left(A + \sum \bar{u}_i B_i\right)^{-1} \sum \bar{u}_i B_i.$$

In many cases a better approximation is obtained by working with the equations in homogeneous form. Consider

$$\frac{d}{dt}\begin{bmatrix} 1 \\ x \end{bmatrix} = \left(\begin{bmatrix} 0 & 0 \\ 0 & A \end{bmatrix} + \sum u_i \begin{bmatrix} 0 & 0 \\ b_i & B_i \end{bmatrix}\right)\begin{bmatrix} 1 \\ x \end{bmatrix}$$

[16] The more recent book of Bloch (2003) covers some of the same material but with more emphasis on the interaction between geometry and control.

and let $\hat{A}$ etc. denote the entries in this equation. As done above, make the change of variables $z = e^{-\hat{A}t}\hat{x}$, This leads to the approximation

$$\hat{x} \approx e^{At}\frac{1}{T}\int_0^T e^{-A\sigma}\sum u_i B_i e^{At} d\sigma \ \hat{x}(0).$$

These ideas were explored in Brockett and Wood (1974) and Wood's thesis (Wood, 1974).

12.4. Quantum control and nuclear magnetic resonance

In the quantum mechanical description of nature provided by the Schrödinger equation, the wave function satisfies a first order differential equation in which the potential energy enters the right-hand side multiplicatively. The concept of force, as it appears in Newton's laws, plays a minor role in quantum mechanics and most efforts directed toward controlling the evolution of the wave function are based on manipulating an appropriate potential energy. The resulting bilinear system is usually infinite dimensional. For example, this is the form of the system when controlling the harmonic oscillator by shifting the location of the minimum of a quadratic potential. Early work by control theorists on quantum systems includes the 1979 work of Butkovskii and Samoilenko (1979) and the work of Haung, Garng, Tarn, and Clark (1983) in 1983. The latter paper quotes extensively from the literature on geometric control. On he other hand, the Bloch model (Bloch, 1946), describing basic nuclear magnetic resonance, is finite dimensional because It describes spin states and there are only a finite number of these.

The angular momentum (spin) of subatomic particles, like the angular momentum of macroscopic bodies can, to some extent, be steered by external forces. Typically, nuclear magnetic resonance measurements are made with the specimen under study being placed in a strong, constant, magnetic field. This creates a statistical equilibrium in which spins aligned with the field are favored over those oppositely aligned. In order to probe the properties of the specimen an input is applied to perturb this equilibrium and then the signals generated as the system returns to its equilibrium state are observed. The mechanism used to perturb the specimen utilizes the fact that any magnetic moment associated with the particle will interact with an external electromagnetic wave. The simplest model for such an interaction is the Bloch model introduced by Felix Bloch in the 1940s. From a control point of view, the model is a bilinear system modeling the evolution of the quantized angular momentum vector, or, more commonly, the evolution of the average value of a large number of angular momentum vectors. The equations of motion for the three components of this average value take the form

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix} = \begin{bmatrix} -\frac{1}{T_1} & u & 0 \\ -u & -\frac{1}{T_2} & \omega \\ 0 & -\omega & -\frac{1}{T_2}\end{bmatrix}\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix}$$

where the T_i account for a relaxation process caused by thermal noise.

For this simple model there is an obvious control strategy which goes back to the earliest days of NMR research. We explain with the help of Fig. 4 which shows a trajectory beginning at the "north pole", being rotated down to the equatorial plane, and then slowly relaxing via a precessing motion drifting back toward the north pole. The initial displacement from the north pole is caused by a short burst of electromagnetic radiation whose frequency is selected to resonate (match) the frequency of the wave function of the quantum system. The subsequent precession of the magnetic moment then generates an electromagnetic wave that can be recorded and analyzed.

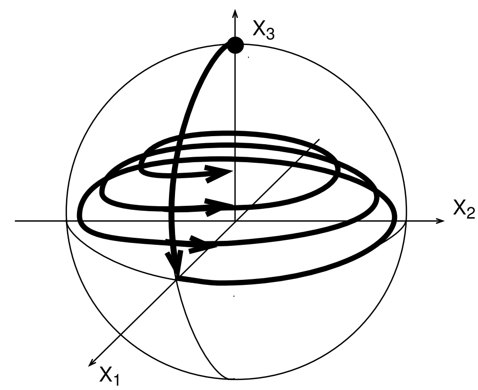

Fig. 4. Controlling the average spin vector.

In spectroscopy the systems of interest are far more complicated than this Bloch model because a number of interacting spins are involved. The strength of the interactions determine the frequency of the resulting signals and these, in turn, provide information about the geometry of the molecular configuration. Beginning with the experimental verification in 1976 of a theoretical idea of Jenner, see Aue, Bartholdi, and Ernst (1978), the much more intricate control of spin systems became popular, leading to what is now known as multi-dimensional Fourier analysis.[17]

References

Abraham, Ralph (with Jerrold Marsden) (1968). *Foundations of mechanics*. Reading, MA: Benjaman/Commings.
Ado, I. D. (1935). Note on the representation of finite continuous groups by means of linear substitutions. *Izvestiya Fizika Matematicheskikh (Kazan)*, *7*, 1–43. Russian language.
Agrachev, A. A., & Gamkrelidze, R. V. (1978). The exponential representation of flows and the chronological calculus. *Mathematics of the USSR-Sbornik*, *107*(4), 467–532.
Airy, G. B. (1840). On the regulator of the clock-work for effecting uniform movement of the equatoreals. *Memoire Royal Astronomical Society*, *11*, 249–267.
Aizerman, M. A., & Gantmcher, F. R. (1963). *Absolute stability of regulator systems*.
Albrecht, F. (1968). Topics in control theory In *Lecture notes in mathematics*: *vol. 63* (2nd ed.). NY: Springer.
Aue, W. P., Bartholdi, E., & Ernst, R. R. (1978). Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *Journal of Chemical Physics*, *64*(1).
Auslander, L., & MacKenzie, G. (1963). *Introduction to differentiable manifolds*. New York: McGraw-Hill.
Baillieul, John (1975). *Some optimization problems in geometric control theory*. (Ph.D. thesis), Harvard University.
Baillieul, John (1981). Controllability and observability of polynomial dynamical systems. *Nonlinear Analysis Theory, Methods and Applications*, *5*(5), 543–552.
Bell, David, & David, Jacobson (1965). *Singular optimal control problems*. Academic Press.
Bellman, Richard (1955). *Matrix theory*. New York: McGraw-Hill.
Beneš, V. E. (1981). Exact finite dimensional filters for certain diffusions with nonlinear drift. *Stochastics*, *5*, 65–92.
Bishop, Richard L., & Richard, J. Crittenden (1964). *Geometry of manifolds*. American Mathematical Society.
Black, Fischer, & Scholes, Myron. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*, 637–654.
Bloch, Felix (1946). Nuclear induction. *Physical Review*, *70*, 460–473.
Bloch, Anthony (2003). Nonholonomic mechanics and control. In *Springer graduate text (with the collaboration of J.Baillieul, P.E. Crouch and J.E. Marsden)*.
Brockett, R. W. (1966). The status of stability theory for deterministic systems. *IEEE Transactions on Automatic Control*, *AC-11*, 596–606.
Brockett, R. W. (1972a). On the algebraic structure of bilinear systems. In R. Mohler, & A. Ruberti (Eds.), *Variable structure control systems* (pp. 153–168). Academic Press.

[17] Optimization problems of this type involving high order bilinear systems have now been studied with considerable success. See, e.g. Khaneja, Roger, and Steffen (2002).

Brockett, R. W. (1972b). System theory on group manifolds and coset spaces. *SIAM Journal on Control*, *10*(2), 265–284.
Brockett, Roger (1973a). Lie algebras and lie groups in control theory. In D. Q. Mayne, & R. W. Brockett (Eds.), *Geometric methods in system theory* (pp. 43–82). Dordrecht, The Netherlands: Reidel Publishing Co.
Brockett, R. W. (1973b). Lie theory and control systems defined on spheres. *SIAM Journal on Applied Mathematics*, *24*(5), 213–225.
Brockett, Roger, & Wood, J. (1974). Electrical networks containing controlled switches. In *Applications of lie group theory to nonlinear network problems, supplement to IEEE international symposium on circuit theory* (pp. 1–11).
Brockett, R. W. (1976a). Nonlinear systems and differential geometry. *IEEE Proceedings*, *64*(2), 61–72.
Brockett, R. W. (1976b). Volterra series and geometric control theory. *Automatica*, *12*, 167–176.
Brockett, R. W. (1976c). Parametrically stochastic linear systems. In R. J. B. Wets (Ed.), *Studies in nonlinear programming*: *vol. 5*. *Stochastic systems: modeling, identification, and optimization* (pp. 8–21). The Netherlands: North Holland Publishers.
Brockett, R. W. (1977). Control theory and analytical mechanics. In C. Martin, & R. Hermann (Eds.), *Geometric control theory* (pp. 1–48). Brookline, Mass: Math. Sci. Press.
Brockett, R. W. (1978). Feedback invariants for nonlinear systems. In *Proceedings of the 1978 IFAC congress, Helsinki, Finland*. Oxford: Pergamon Press.
Brockett, R. W., & Clark, J. M. C. (1980). The geometry of the conditional density equations. In O. L. R. Jacobs, et al. (Eds.), *Analysis and optimization of stochastic systems* (pp. 299–309). New York: Academic Press.
Brockett, R. W. (1980). Some remarks on finite dimensional nonlinear estimation. *Astérique, Société de Mathématique de France*, *75/76*, 47–55.
Brockett, R. W. (1981a). Nonlinear systems and nonlinear estimation theory. In M. Hazewinkel, & J. C. Willems (Eds.), *Stochastic systems* (pp. 441–477). Dordrecht, The Netherlands: Reidel Publishing Co.
Brockett, Roger (1981b). Control theory and singular riemannian geometry. In P. Hilton, & G. Young (Eds.), *New directions in applied mathematics* (pp. 11–27). New York: Springer-Verlag.
Brockett, Roger (1983). Asymptotic stability and feedback stabilization. In R. W. Brockett, et al. (Eds.), *Progress in mathematics*: *vol. 27*. *Differential geometrics in control theory* (pp. 181–191). Boston, MA: Birkhäuser Publishers.
Brockett, R. W. (1984). Nonlinear control theory and differential geometry. In *Proceedings of the international congress of mathematicians* (pp. 1357–1367). Warszawa: Polish Scientific Publishers.
Bruni, C., Di Pillo, G., & Koch, G. (1971). On the mathematical models of bilinear systems. *Ricerche di Automatica*, *2*, 11–26.
Bruni, C., Di Pillo, G., & Koch, G. (1974). 'Bilinear systems: an appealing class of nearly linear' systems in theory and applications. *IEEE Transactions on Automatic Control*, *AC-19*, 334–348.
Brunovsky, P. (1970). A classification of controllable linear systems. *Kybernetlka*, *3*, 173–187.
Butkovskii, A. G., & Samoilenko, Y. I. (1979). Control of quantum systems. *Automation and Remote Control*, *40*, 485–502. 629-645.
Byrnes, C. I., & Krener, A. J. (1983). On the existence of globally (f, g)-invariant distributions. In R. W. Brockett, et al. (Eds.), *Progress in mathematics*: *27*. *Differential geometrics in control theory* (pp. 209–225). Boston, MA: Birkhäuser Publishers.
Carleman, T. (1932). Application de la théorie des équations intégrales singularees aux equations différentielles de la dynamique. *Transaction of Arkiv for Matematik Astrometriya Fysikkens*, *22B*.
Chen, K. T. (1957). Integration of paths, Geometric invariants and a Generalized Baker-Hausdorff formula. *Annals of Mathematics*, *65*(1).
Chen, K. T. (1962). Decomposition of differential equations. *Mathematische Annalen*, *146*, 263–278.
Chow, W. L. (1939). Uber systeme von linearean partiellen differentialgleichungen erster ordinung. *Mathematische Annalen*, *117*, 98–105.
Clark, J. M. C. (1973). An introduction to stochastic differential equations. In D. Q. Mayne, & R. W. Brockett (Eds.), *Geometric methods in system theory* (pp. 131–149). Dordrecht, The Netherlands: Reidel Publishing Co.
Crouch, P. E. (1977). Dynamical realizations of finite Volterra series. *SIAM Journal on Control and Optimization*, *19*, 177–202.
Crouch, P. E., & Bonnard, B. (1980). *An appraisal of linear analytic systems theory with applications to attitude control*. European Space Agency.
D'Allesandro, P., Isidori, A., & Ruberti, A. (1972). Realization and structure theory of bilinear dynamical systems. In *Inst. di Automatica, University of Rome, Rep. 2-04*. *SIAM Journal on Control*, *12*, 517–535.
Dubins, L. E. (1957). On curves of minimal length with a constraint on average curvature and with prescribed initial and terminal positions and tangents. *American Journal of Mathematics*, *79*, 497–516.
Elliott, David L. (1971). A consequence of controllability. *Journal of Differential Equations*, *10*, 364–370.
Flanders, H. (1963). *Differential forms*. New York: Academic Press.
Fliess, M., & Lamnabhi-Lagarrigue, F. (1982). Application of a new functional expansion to the cubic anharmonic oscillator. *Journal of Mathematical Physics*, *23*, 495–502.
Fliess, M., & Normand-Cyrot, D. (1982). 'Alg'ebres de Lie nilpotentes, formule de Baker-Campbell-Hausdorff et intégrales itérées de K.T. Chen'. In *LNM*: *vol. 920*. *Séminaire de Probabilités*. Springer-Verlag.
Fliess, M., Lévine, J. L., Martin, P., & Rouchon, P. (1995). Flatness and defect of non-linear systems: introductory theory and examples. *International Journal of Control*, *61*, 1327–1361.
Fisk, D. L. (1963). *Quasi-martingales and stochastic integrals, tech. report no. 1*. Dept. Math. Mich. State Univ.
Freund, E. (1975). The structure of decoupled non-linear systems. *International Journal of Control*, *21*, 443–450.
Gantmacher, R. F. (1959). *Theory of matrices*. New York: Chelsea.
Gabasov, R., & Kirillova, F. M. (1972). High order necessary conditions for optimality. *SIAM Journal on Control*, *10*, 127–168.
Gaveau, B. (1977). Principe de Moindre Action, Propagation De la Chaleur et Estimées Sous Elliptiques sur certains Groupes Nilpotents. *Acta Mathematica*, *139*.
Griffiths, P. A. (1983). *Exterior differential systems and the calculus of variations*. Boston: Birkerhauser.
Halmos, P. R. (1958). *Finite dimensional vector spaces*. Princeton, NJ: Van Nostrand.
Hahn, W. (1967). *Stability of motion*. Berlin: Springer-Verlag.
Haung, Garng, M., Tarn, T. J., & Clark, John W. (1983). On the controllability of quantum mechanical systems. *Journal of Mathematical Physics*, *24*, 2608–2618.
Haynes, G. W., & Hermes, H. (1970). Nonlinear controllability via Lie theory. *SIAM Journal on Control*, *8*, 450–460.
Hurbert, C. (1981). Spacecraft attitude acquisition from an arbitrary spinning or tumbling state. *Journal of Guidance and Control*, *4*, 164–170.
Hermes, H., & Haynes, G. W. (1963). On the nonlinear control problem with the control appearing linearly. *SIAM Journal on Control*, *1*, 85–108.
Hermes, H. (1974). On local and global controllability. *SIAM Journal on Control and Optimization*, *12*, 252–261.
Hermes, H. (1976). Local controllability and sufficient conditions in singular problems. *Journal of Differential Equations*, *20*, 213–232.
Hermes, H. (1978). Lie algebras of vector fields and local approximations of attainable sets. *SIAM Journal on Control and Optimization*, *33*(16), 715–727.
Hermann, R. (1963). On the accessibility problem in control theory. In J. P. LaSalle, & S. Lefschiftz (Eds.), *Nonlinear differential equations and nonlinear mechanics*. New York: Academic Press.
Hirschorn, Ronald (1973). Topological semigroups, sets of generators, and controllability. *Duke Mathematical Journal*, *40*(4), 937–947.
Hirschorn, Ronald (1981). (A,B)-Invariant distributions and disturbance decoupling of nonlinear systems. *Journal on Control and Optimization*, *19*, 1–19.
Hooker, W. W., & Margoulis, G. (1965). The dynamical attitude equations for an n-body satellite. *Journal of the Astronautical Science*, *12*, 123–128.
Hormander, L. (1967). Hypoelliptic second order differential equations. *Acta Mathematic*, *119*, 147–171.
Isidori, A., Krener, A. J., Claudio, Gorgi-Georgi, & Monico, S. (1981). Nonlinear decoupling via feedback: a differential geometric approach. *IEEE Transactions on Automatic Control*, *AC-26*, 331–345.
Isidori, A., Krener, A. J., Claudio, Gorgi-Georgi, & Monico, S. (1981). Locally f, g-invariant distributions. *Systems & Control Letters*, *1*, 12–15.
Itô, K. (1946). On a stochastic integral equation. *Proceedings of the Impresa Academie*, *22*(2), 1946.
Jakubczyk, B., & Respondek, W. (1980). On linearization of control systems. *Bulletin de l'Academie Polonaise des Sciences Matematica*, *28*, 517–522.
Jakubczyk, B. (1980). Existence and uniqueness of realizations of non-linear systems. *SIAM Journal on Control and Optimization*, *18*, 455–471.
Kalman, R. E. (1960a). On the structural properties of linear constant multivariable systems. In *Proceedings of the 1966 IFAC conference*. London: Butterworths.
Kalman, R. E. (1960b). Contributions to the theory of optimal control. *Sociedad MatemáticaMexicana.Boletín*, *5*, 102–119.
Kalman, R. E., & Bertram, J. E. (1960). Control system analysis and design via the "second method" of Lyapunov. I. Continuous-time systems. *Transactions of the ASME. Series D, Journal of Basic Engineering*, *82*, 371–393.
Kalman, R. E. (1963). Mathematical description of linear dynamical systems. *SIAM Journal on Control*, *1*(2), 152–192.
Kalman, R. E. (1971). Kronecker invariants and feedback. In *Proceedings of the conference on ordinary differential equations*. Mathematical Research Center, Naval Research Lab.
Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *ASME Journal of Basic Engineering*, *83*, 95–108.
Kane, T. R., & Scher, M. P. (1969). A dynamical explanation of the falling cat phenomena. *International Journal of Solids and Structures*, *5*, 663–670.
Kelley, H. J., Kopp, R. E., & Moyer, H. G. (1966). Singular extremals. In G. Leitmann (Ed.), *Topics in optimization, Vol. II*. New York: Academic Press, Chapter 2.
Khaneja, N., Roger, Brockett, & Steffen, Glasser (2002). Sub-Riemannian geometry and time optimal control of three spin systems: quantum gates and coherence transfer. *Physical ReviewA*, *65*, 032301.
Kokotovic, P. V. (1992). The joy of feedback: nonlinear and adaptive. *IEEE Control Systems Magazine*, *12*(3), 7–17.
Kossakowski, A. (1972). On quantum statistical mechanics of non-Hamiltonian systems. *Reports on Mathematical Physics*, *3*, 247–274.
Krasnoselskii, M. A., & Zabreko, P. P. (1984). *Geometrical methods of nonlinear analysis*. Berlin: Springer-Verlag.
Krasovskii, N. N. (1963). *Problems of the theory of stability of motion*. Stanford, CA: Stanford University Press, (in Russian).
Krener, A. J. (1971). *A generalization of the pontryagin maximal principle and the bang–bang principle*. (Ph.D. thesis). Berkeley: Dept. of Mathematics, University of California.
Krener, A. J. (1973). On the equivalence of control systems and the linearization of nonlinear systems. *SIAM Journal on Control*, *11*(4), 670–676.
Krener, A. J. (1974). A generalization of Chow's theorem and the bang–bang theorem to Nonlinear control problems. *SIAM Journal on Control*, *12*(4), 43–52.
Krener, A. J. (1975). Bilinear and Nonlinear realizations of input–output maps. *SIAM Journal on Control*, *13*(4), 827–834.

Krener, A. J. (1977). The high order maximum principle and its application to singular extremals. *SIAM Journal on Control*, *15*(2), 256–293.
Krener, A. J. (1983). Linearization by output injection and nonlinear observers. *Systems & Control Letters*, *3*(1), 47–52.
Krishnaprassad, P. S. (1985). Lie-Poisson structures, dual spin spacecraft and asymptotic stability. *Nonlinear Analysis, Theory, Methods and Applications*, *9*(10), 1011–1035.
Kučera, Jan (1966). Solution in large of control problem $\dot{x} = (A(1-u) + Bu)x$. *Czechoslovak Mathematical Journal*, *16*, 600–693.
Kučera, Jan (1967). Solution in large of control problem $\dot{x} = (Au+Bv)x$. *Czechoslovak Mathematical Journal*, *17*, 91–96.
Kurzweil, J. (1956). On the inversion of Liapunov's second theorem on stability of motion. *AMS Translation Series II*, *2*, 19–77.
Lawden, D. F. (1962). Optimal intermediate-thrust arcs in a gravitational field. *Astronautica Acta*, *8*, 106–123.
LaSalle, J. P. (1960). Some extensions of Liapunov's second method. *IRE Transactions on Circuit Theory*, *CT-7*, 520–527.
Lefschetz, Solomon (1957). *Differential equations: geometric theory*. Interscience.
Lesiak, C. M., & Krener, A. (1978). Existence and Uniqueness of Volterra Series. *IEEE Transactions on Automatic Control*, *TAC 78*.
Likins, P. W. (1967). Attitude stability criteria for dual spin spacecraft. *Journal of Spacecraft and Rockets*, *4*, 1638–1643.
Lindblad, G. (1976). On the generators of quantum dynamical semigroups. *Acta Mathematic*, *48*, 119–130.
Lobry, C. (1973). Dynamical polysystems and control theory. In D. Q. Mayne, & R. W. Brockett (Eds.), *Geometric methods in system theory* (pp. 1–42). Dordrecht, The Netherlands: Reidel Publishing Co.
Lobry, C. (1974). Controllability of nonlinear systems on compact manifolds. *SIAM Journal on Control*, *12*, 1–4.
Lur'e, A. I., & Postnikov, V. N. (1944). On the theory of stability of control systems. *Applied mathematics and mechanics*, *44*, in Russian.
Magnus, W. (1954). On the exponential solution of differential equations for a linear operator. *Communications on Pure and Applied Mathematics*, *7*, 1954.
Malkin, I. G. (1959). *Theory of the stability of motion*. Washington DC: United States Atomic Energy Commission.
Markus, Larry, & Ernest, Bruce Lee (1967). *SIAM series in applied mathematics*, *Foundations of optimal control theory*. Wiley.
Maxwell, J. C. (1868). On governors. *Proceedingsof the Royal Society*, *16*, 270–283.
Mayne, D. Q., & Brockett, Roger (1973). *Geometric methods in system theory*. Reidel: Dordrecht, The Netherlands.
McKean, H. P. (1969). *Stochastic integrals*. New York: Academic Press.
Merton, Robert. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, *4*, 141–183.
Massera, J. L. (1956a). Contributions to stability theory. *Annals of Mathematics*, *64*, 182–206.
Meyer, G. 1966. On the use of Euler's Theorem on Rotations for the synthesis of attitude control systems, NASA Technical Note, TND-3643, Ames Research Center, Moffett Field, California.
Meyer, G., & Cicolani, L. 1975. A Formal Structure for Advanced Automatic Flight-Control Systems, NASA Technical Note, TND-7940, Ames Research Center, Moffett Field, California.
Milnor, John W. (1956). On manifolds homeomorphic to the 7-sphere. *Annals of Mathematics*, *64*(2), 399–405.
Middlebrook, R. D., & C'uk, S. (1977). Modeling and analysis methods for DC-to-DC switching converters. *IEEE International Semiconductor Power Converter Conference*,.
Mortensen, Richard (1968). A globally stable linear attitude regulator. *International Journal of Control*, *8*(3), 297–302.
Neimark, Ju. I., & Fufaev, N. A. (1972). *Translations of mathematical monographs*: *vol. 33. Dynamics of nonholonomic systems*. Providence, RI: AMS.
Newton, George C. (with Leonard Gould and James Kaiser), Jr. (1957). *Analytical design of linear feedback controls with chart in rear pocket*. New York: John Wiley.
Poincaré, Henri (1895). Analysis situs. *Journal de l École Polytechnique*, *1*, 1–123.
Pontryagin, L. S, Boltyanskii, V. G., Gamkrelidze, R. V., & Mishchenko, E. F. (1959). *The mathematical theory of optimal processes*. Interscience. English translation, Russian.
Popov, V. M. (1962). Absolute stability of nonlinear systems of automatic control. *Automation and Remote Control*, *22*, 857–875. Russian original August 1961.
Rashevskii, P. K. (1938). About connecting two points of complete non-holonomic space by admissible curve. *Uchebnykh Zapiski Pedagogicheskogo Institute of Libknexta*, *2*, 83–94. (in Russian).
Riemann, Bernhard 1868. Über ber die Hypothesen welche der Geometrie zu Grunde liegen, (1954 Habilitationsschrift, Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen.).
Rink, R. E., & Mohler, R. R. (1962). Completely controllable bilinear systems. *SIAM Journal on Control*, *6*(3).
Roberson, R. E., & Wittenburg, J. (1968). A dynamical formalism for an arbitrary number of interconnected rigid bodies with reference to the problem of satellite attitude control. In *Proceedings 3rd IFAC congress, London*.
Rosenbrock, H. H. (1970). *State space and multivariable theory*. London: Nelsons.
Rossmann, Wulf (2003). *Oxford graduate texts in mathematics*, *Lie groups: an introduction through linear groups*. Oxford University Press.
Sanders, Seth R., et al. (1991). Generalized averaging method for power conversion circuits. *IEEE Transactions on Power Electronics*, *6*, 251–259.
Sternberg, Shlomo (1964). *Lectures on differential geometry*. Englewood Cliffs, NJ: Prentice Hall.
Stratonovich, Ruslan (1963). *Topics in the theory of random noise, vol. 2*. London: Gordon and Breach.
Strichartz, Robert S. (1986). Sub-rRemannian Geometry. *Journal of Differential Geometry*, *24*, 221–263.
Sussmann, H. J., & Jurjdevic, V. (1972). Controllability of nonlinear systems. *Journal of Differential Equations*, *12*, 313–329.
Sussmann, H. J. (1972). The Bang-Bang Problem for Linear Control Systems in $Gl(n)$. *SIAM Journal on Control*, *10*(3), 470–476.
Sussmann, H. J. (1973). Orbits of families of vector fields and integrability of distributions. *Transactions of the American Mathematical Society*, *180*, 171–188.
Sussmann, H. J. (1977). Existence and uniqueness of minimal realizations of nonlinear systems. *Mathematical Systems Theory*, *10*, 263–284.
Sussmann, H. J. (1978). A sufficient condition for local controllability. *SIAM Journal on Control and Optimization*, *16*(5), 790–802.
Wei, J., & Norman, E. (1963). Lie algebraic solutions of linear linear differential equations. *Journal of Mathematical Physics*, *4*, 575–581.
Wei, J., & Norman, E. (1964). On the global representation of the solutions of linear differential equations as a product of exponentials. *Proceedingsof the American Mathematical Society*, *15*, 327–334.
Whittaker, E. T. (1927). *A treatise on the analytical dynamics of particles and rigid bodies*. Cambridge, England: Cambridge University Press.
Wilson, F. W. (1967). The structure of the level surfaces of a liapunov function. *Journal of Differential Equations*, *4*, 323–329.
Wonham, W. M. (1965). Some applications of stochastic differential equations to optimal nonlinear filtering. *SIAM Journal on Control*, *2*, 347.
Wonham, W. M., & Morse, A. S. (1972). Feedback invariants of linear multivariable systems. *Automatica*, *8*, 93–100.
Wood, Jonathan (1974). *Power conversion in electrical networks*. (Ph.D. thesis). Dept. of Electrical Engineering, MIT.
Yakubovich, V. A. (1962). Solution of certain matrix inequalities in the stability theory of nonlinear control systems. *Doklady Akademii Nauk SSSR*, *143*, 1304–1307. English translation in *Soviet Math.Dokl.***3**, pp. 620–623.
Zubov, V. (1964). *Methods of A.M. Liapunov and their applications*. Amsterdam: Noordhoff.

Roger Brockett is the An Wang Research Professor of Electrical Engineering and Computer Science in the School of Engineering and Applied Sciences at Harvard University. He received his B.S., M.S. and Ph.D. degrees from Case Western Reserve University and taught for six years in the Electrical Engineering Department at MIT before joining the Harvard faculty in 1969. He retired from the teaching faculty in 2012. He has contributed to the theory of automatic control with work on stability, nonlinear control, feedback linearization, aspects of system identification, nonlinear estimation, pole placement, hybrid systems, and robotics. More recently his work has involved problems arising in the study of intelligent machines. Areas of particular interest include the problem of motion control, minimum attention control, synchronization and the investigation of new paradigms in quantum control. Professor Brockett has been recognized with awards from the American Automatic Control Council, IEEE, ASME and SIAM for his contributions to research and education. He is a fellow of the IEEE, SIAM, and AMS and is a member of the US National Academy of Engineering.