AMS/IP Studies in Advanced Mathematics, vol 39, 2007, pp 23-35.

Optimal control of the Liouville equation

R. W. Brockett

Abstract. We explore the role of a special class of optimization problems involving controlled Liouville equations. We present some new results on the controllability of the Liouville equation and discuss the optimal control of its moments for some important special cases. The problems treated suggest a new domain of applicability for optimal control, one which can include constraints more appropriate for the synthesis of control systems for a large class of real world systems.

1. Introduction

In this paper we argue that some of the important limitations standing in the way of the wider use of optimal control can be circumvented by explicitly acknowledging that in most situations the apparatus implementing the control policy will judged on its ability to cope with a distribution of initial states, rather than a single state. This makes it appropriate to represent the system dynamics in terms of its Liouville equation and to formulate performance measures in that setting. That is, we argue in favor of replacing the usual control model $\dot{x} = f(x,u)$ by the first order partial differential equation

$$\frac{\partial \rho(t,x)}{\partial t} = -\left\langle \frac{\partial}{\partial x}, f(x,u)\rho(t,x) \right\rangle$$

and for the consideration of performance measures which include *non trajectory dependent terms* such as the second and third terms on the right-hand side of

$$\eta = \int_0^T \int_S \rho(t,x)L(x,u)dxdt + \int_S \left(\frac{\partial u}{\partial x}\right)^2 dx + \int_0^T \left(\frac{\partial u}{\partial t}\right)^2 dt$$

Later on we give a number of examples to support our contention but, in brief, we may say that by forcing the specification of a range of initial conditions this formulation allows one to include some important types of performance measures not expressible in standard deterministic optimal control theory. In specific cases this allows the designer to express, in mathematical terms, issues related to the role of feedback, robustness and cost of implementation. This approach has some aspects in common with stochastic control but occupies a position of intermediate complexity, lying between deterministic calculus of variations on finite dimensional manifolds and, the usually much less tractable, control of finite dimensional diffusion processes.

In almost all applications of automatic control there is a trade off between the gains to be had from trajectory optimization and the cost of the apparatus required to implement the optimal policy. A simpler apparatus that generates commands

This work was supported in part by the US Army Research Office under Boston University prime GC169369 NGD, the National Science Foundation grant number ECS 0424342, and by DARPA under Stanford University Prime PY-1606.

that closely approximate the optimal trajectories but costs less to build and maintain can be preferable to a more complex apparatus that precisely generates the optimal trajectories. Such considerations apply both to technological applications and to the study of motion control in animals where evolutionary pressure seeks to optimize the allocation of neural circuitry. In spite of these arguments, there is almost nothing in the control literature attempting to translate these ideas into mathematics. In this paper we formulate and solve some problems relating to such questions.

We begin with some background. Typically problems in deterministic optimal control are formulated in terms of a variable x, called the *state*, related to a variable u, called the *control*, through a differential equation, $\dot{x} = f(x,u)$ and an initial equation $x(0) = x_0$, together with a performance measure

$$\eta = \int_0^T L(x,u)dt$$

The problem is to find a control $u^*(x,t)$ that minimizes η. Assuming that an optimal control u^* exists, it might be expressed as a function of t and x_0 and otherwise independent of x or it might be expressed in terms of t and $x(t)$. Engineers express the matter by saying that when the optimal u is expressed as a function of t and x_0 it is in *open loop* form, and when it is expressed in terms of $x(t)$, it is expressed in *closed loop* form. The distinction between open loop and closed loop is of little significance for the mathematical problem stated because the trajectories are the same in either case, but the difference is quite significant when it comes to using such results. In the first place, some apparatus must be made available to generate the function u and the cost and complexity of this apparatus can be expected to depend on this distinction. Secondly, the mathematical model for the system and the characterization of the initial state is inevitably inexact and the effect of the error will depend on whether u is realized in closed loop form or open loop form.

The situation found in stochastic control is different. In this subject the distinction between open loop performance and closed loop performance is quite obvious. It is standard to formulate problems in terms of a probabilistic model for the observations and to restrict the control to be a *past measurable function* of the observables. In this way the distinction between open loop and closed loop control is woven into the problem formulation from the beginning. Typically problems are described in terms of a stochastic differential equation such as $dx = f(x,u)dt + gdw$ and a vector of observables $dy = h(x)dt + d\nu$. The objective is to minimize the expected value of some functional. In this case there is a difference between the optimal performance obtainable when u is constrained to be a function of time and the statistics of $x(0)$ only, versus the performance obtainable when it is allowed to depend on the value of x at the present time. Typically no open loop control can perform as well as the closed loop one. One concrete aspect of this distinction is is illustrated by the fact that in the Fokker-Planck equation for the probability density of x, namely

$$\frac{\partial \rho(t,x)}{\partial t} = -\left\langle \frac{\partial}{\partial x}, f(x,u)\rho(t,x) \right\rangle + \frac{1}{2}\left\langle \frac{\partial^2}{\partial x^2}, gg^T\rho(t,x) \right\rangle \quad ; \quad \rho(0,x) = \text{given}$$

u is acted upon by partial derivative with respect to x so that the type of functional dependence matters. However, the use of stochastic control theory requires considerable effort to gather the statistical data needed to define the problem, and even then is often quite difficult to solve.

In this paper we describe an alternative formulation of control problems that attempts to capture more fully the implications of the various ways of expressing u without encountering the full range of difficulties associated with stochastic control. As in our earlier work [1], this approach involves controlling a Liouville equation and leads to a class of variational problems that seem to have been largely ignored until now.

2. Controlling the Liouville Equation

Let X be an oriented differentiable manifold with a fixed, nondegenerate, volume form dv. Let $\phi : X \to X$ be a diffeomorphism. If ρdv is a nonnegative density on X and if ϕ is orientation preserving, then ϕ acts on densities according to

$$\rho(\cdot) \mapsto \rho(\phi^{-1}(\cdot))/\mathrm{detJ}_\phi$$

where J_ϕ is the Jacobian of ϕ. In this sense $\mathrm{Diff}_O(X)$, the set of orientation preserving diffeomorphisms, generates an orbit through a given ρ. It is known, for example, that if the manifold is compact then this action acts transitively on the strictly positive densities. (See Moser [2].) If the densities are only assumed to be nonnegative this action is clearly not transitive.

Associated with a control equation of the form $\dot{x} = f(x, u)$ there is a first order partial differential equation containing a control term which governs the evolution of smooth densities,

$$\frac{\partial \rho(t,x)}{\partial t} = -\left\langle \frac{\partial}{\partial x}, f(x,u)\rho(t,x) \right\rangle$$

This is a control dependent version of the Liouville equation of mathematical physics. If the underlying manifold is $\mathbb{R}^n$ we can sketch its derivation as follows. Assume that ρ is a smooth function and that $\psi(x)$ is a smooth test function with compact support. On one hand, the time evolution of the average value of ψ is given by

$$\frac{\partial}{\partial t}\int_{\mathbb{R}^n} \psi(x)\rho(t,x)dx = \int_{\mathbb{R}^n} \psi(x)\frac{\partial}{\partial t}\rho(t,x)dx$$

on the other hand,

$$\frac{d}{dt}\psi(x) = \left\langle \frac{\partial \psi}{\partial x}, f(x,u) \right\rangle$$

and so

$$\frac{\partial}{\partial t}\int_{\mathbb{R}^n} \psi(x)\rho(t,x)dx = \int_{\mathbb{R}^n} \rho(t,x)\left\langle \frac{\partial \psi(t,x)}{\partial x}, f(x,u) \right\rangle dx$$

An integration-by-parts then shows that for all such ψ

$$\int_{\mathbb{R}^n} \psi(x)\frac{\partial}{\partial t}\rho(t,x)dx = -\int_{\mathbb{R}^n} \psi(x)\left\langle \frac{\partial \rho(t,x)}{\partial x}, f(x,u) \right\rangle dx$$

In view of the fact that ψ is arbitrary we see that a smooth ρ evolves as given.

If $\rho(0,x)$ is nonnegative with $\int \rho(0,x)dx = 1$, then the Liouville equation can be thought of as the evolution equation for an initial probability density of x under the flow defined by the deterministic equation $\dot{x} = f(x,u)$. As is well known, the solution of the Liouville equation can be expressed in terms of the general solution of $\dot{x} = f(x, u)$. If the solution of $\dot{x} = f(x, u(x))$ is such that the initial value x_0 goes to $\phi(t, x_0)$ at time t then the solution of the Liouville equation is

$$\rho(t, x) = \rho(0, \phi^{-1}(t, x))/\mathrm{detJ}_\phi(\mathrm{x})$$

where ϕ^{-1} denotes the result of solving $x = \phi(t, x_0)$ for x_0 and J is the Jacobian of the map ϕ. For example, if $\dot{x} = x$ the map is $\phi(t, x) = e^t x$ and the initial density is $\pi^{-1}/(1 + x^2)$ then the density at time t is $e^{-2t}\pi^{-1}/(1 + e^{-2t}x^2)$.

Given $\dot{x} = f(x, u)$ together with a performance measure $\eta = \int_0^T L(x, u)dt$ such that an optimal control exists, we have an optimal value function

$$\eta^*(x(0)) = \min_u \int_0^T L(x, u)dt$$

If S is the support of the initial density We can evaluate the average performance of the optimal system using

$$\mathcal{E}\eta(x) = \int_S \rho(0, x)\eta^*(x)dx$$

If we place no further conditions on the way u depends on x and t then the Liouville point of view contributes nothing essentially new because

$$\Gamma^*(\rho(0, x)) = \min_u \int_0^T \int_S \rho(t, x)L(x, u)dxdt$$

That is, $\Gamma^*(\rho(0, x)) = \mathcal{E}\eta^*$. However, if we place conditions on u with the goal of making the system more robust with respect to errors or with the goal of simplifying the apparatus necessary to generate the control, such as requiring that u be affine in x or limiting the norm of $\partial u/\partial x$, etc. then the most we can say is that

$$\Gamma(\rho(0, x)) \geq \mathcal{E}\eta(x)$$

and we have a genuinely new class of problems.

To explore these issues we investigate the control of the Liouville equation with performance measures of the type suggested in the introduction. We consider three examples.

Example 1: Consider the scalar equation $\dot{x} = u$ with associated Liouville equation

$$\frac{\partial \rho(t, x)}{\partial t} = -\frac{\partial}{\partial x} u(t, x)\rho(t, x)$$

Suppose we constrain u to be affine function of x and want to minimize

$$\Gamma = \int_0^T \int_S \rho(t, x)x^2 dxdt + \int_0^T \left(\frac{\partial u}{\partial x}\right)^2 + \left(u - \frac{\partial u}{\partial x}x\right)^2 dt$$

The terms in Γ can be interpreted as follows. The ρx^2 term penalizes trajectories for deviating from $x = 0$. The $\partial u/\partial x$ term penalizes the design for using high gain feedback whereas the remaining term penalizes the control for using large open loop signal. According to standard linear quadratic control theory, the minimum value of

$$\eta = \int_0^T x^2 + v^2 dt$$

for a system $\dot{x} = a(t)x + v$ is given by $\eta^* = k(0)x^2(0)$ where

$$\dot{k} = 2ak + 1 - k^2 \;\; ; \;\; k(T) = 0$$

Thus our problem can be recast as that of minimizing

$$\Gamma = k(0)\int_{\mathbb{R}} \rho(0,x)x^2dx + \int_0^T a^2dt$$

Solving for a in terms of k and denoting the second moment of x at time 0 as m, the problem is to minimize

$$\Gamma = mk(0) + \int_0^T \left(\frac{\dot{k}-k^2+1}{2k}\right)^2 dt$$

Letting $r = \ln k$ the integrand becomes $(\dot{r}/2 - \sinh r)^2$. Applying Euler-Lagrange theory we see that $\ddot{r} - 4\cosh r \sinh r = 0$. Thus there exists a constant d such that

$$\dot{r} = \sqrt{d - 4\sinh^2 r}$$

so that $k = e^r$ and $a = (\dot{k} - k^2 + 1)/2k$.

Example 2: Consider the same system as above and again restrict the control law to be affine in x. Now we consider the least squares control of the mean and variance of the density. This class of problems will be discussed in more detail in section 5. The system is

$$\dot{x} = -kx + v$$
$$\dot{\sigma} = -2k\sigma$$

and we want to find k and v so as to steer the mean and variance to desired values while minimizing

$$\eta = \int_0^T k^2 + v^2dt$$

In anticipation of using the Pontryagin maximum principle, we form the Hamiltonian

$$h = -pkx + pv - 2qk\sigma + v^2/2 + k^2$$

Thus $v_{\text{opt}} = -p$ and $k_{\text{opt}} = q\sigma + px/2$ and

$$\dot{p} = kp$$
$$\dot{q} = 2kq$$

¿From this we see that $q\sigma$ is a constant and thus v and k satisfy

$$\dot{k} = -v^2/2$$
$$\dot{v} = kv$$

This implies that $\ddot{k} = -2\dot{k}k$ and so $\dot{k} = -k^2 + c$ and

$$k(t) = \sqrt{c}\tanh(\sqrt{c}\,t + \alpha)\ \ ;\ \ v = \sqrt{2c - 2k^2}$$

Example 3: Again we consider the scalar control problem $\dot{x} = u$ with the distribution of initial conditions being given by a density $\rho_0(x)$. The Liouville equation is as above but now we consider the minimization of

$$\eta = \int_0^\infty \int_{-1}^1 \rho(t,x)ax^2dxdt + \int_{\mathbb{R}} \left(\frac{\partial u}{\partial x}\right)^2 dx$$

If we constrain $u(x)$ to be a function of x alone and if $\int_0^\infty x^2dt$ is finite then

$$\int_0^\infty x^2dt = \int_0^{x(0)} \frac{x^2}{u(x)}dx \text{ and } \int_0^\infty \mathrm{u^2dt} = \int_0^{\mathrm{x(0)}} \mathrm{u(x)dx}$$

so that

$$\int_0^\infty \int_{\mathbb{R}} \rho(t,x) a x^2 dx dt = \int_{\mathbb{R}} \rho_0(x_0) \left(\int_0^{x(0)} a \frac{x^2 dx}{u(x)} + bu(x) \right) dx_0$$

Thus the functional to be minimized can be written as

$$\eta = \int_{\mathbb{R}} \rho_0(x) \left(\int_0^x a \frac{w^2}{u(w)} dw \right) dx + \int_{\mathbb{R}} \left(\frac{\partial u}{\partial x} \right)^2 dt dx$$

In the special case where ρ_0 is a delta function centered at x_0 the term involving ρ_0 can be simplified giving

$$\eta = \int_0^{(x(0)} \left(a \frac{x^2}{u(w)} \right) dx + \int_{\mathbb{R}} \left(\frac{\partial u}{\partial x} \right)^2 dt dx$$

The indicated partial derivative is actually a total derivative and so an application of the Euler-Lagrange operator gives

$$\frac{d^2 u}{dx^2} + a \frac{x^2}{u^2} = 0$$

Because the solution x should remain at zero when reaching zero, it is necessary to impose the condition $u(0) = 0$ on solutions of the Euler Lagrange equation. It is easy to verify that the appropriate solution is

$$u = -(9a/4)^{1/3} x^{4/3}$$

In this case the integral squared error is proportional to $x^{5/2}(0)$.

3. Controllability of the Liouville Equation

Before considering optimal control problems involving the Liouville equation further we establish some results describing the extent to which one can influence the evolution of a density through the use of feedback control. With this in mind we now establish a controllability property of the Liouville equation associated with controllable linear systems.

Consider a system evolving in $\mathbb{R}^n$ described by

$$\dot{x} = Ax + Bu$$

The corresponding Liouville equation is

$$\frac{\partial \rho(t,x)}{\partial t} = -\left\langle \frac{\partial}{\partial x}, (Ax + Bu)\, \rho(t,x) \right\rangle$$

A pair (A, B) with A and n by n matrix and B an n by m matrix is said to be *controllable* if the columns of $B, AB, ..., A^{n-1}B$ contains a basis for $\mathbb{R}^n$. For such systems we introduce a one parameter family of symmetric, nonnegative definite matrices

$$W(0,t) = \int_0^t e^{-A\tau} B B^T e^{-A^T \tau} d\tau$$

The controllability property implies that $W(0,t)$ is nonsingular for positive t. From the variation of constants formula we see that the solution x corresponding to the control

$$u(t) = B^T e^{-A^T t} W^{-1}(0,T)(e^{-At} x_f - x_0)$$

is

$$x(t) = e^{At}\left(x_0 + \int_0^t e^{-A\sigma}BB^Te^{-A^T\sigma}d\sigma W^{-1}(0,T)(e^{-AT}x_f - x_0)\right)$$

steers the system from $x(0) = x_0$ to $X(T) = x_f$

As noted above, it is not true that any initial density can be transformed into any other density at a later time. For example, open sets on which the initial density vanishes will necessarily map into open sets on which subsequent densities vanish. However, it is possible to reach from $\rho_0(\cdot)$ a large class of densities.

We begin with a useful lemma. We use the notation $||M||$ to denote the induced Euclidean norm of a matrix.

Lemma 1: Let $\dot{x} = Ax + Bu$ be a controllable linear system. Let $\psi : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism with Jacobian $J_\psi(x)$. Suppose $T > 0$ is given and that for some $k < 1$ and all x, y the map ψ satisfies the Lipschitz inequality

$$||e^{-AT}\psi(x) - x - e^{-AT}\psi(y) + y|| \leq k||x - y||$$

Then there exists a control law $u(x,t)$ that steers the Liouville equation from the given initial condition $\rho(0,x) = \rho_0(x)$ to $\rho(T,x) = \rho_0(\psi(x)))/\mathrm{det}(\mathrm{J}_\psi(\mathrm{x}))$.

Proof: The solution of $\dot{x} = Ax + Bu$ with the control

$$u(t,x_0) = B^Te^{-A^Tt}W^{-1}(0,T)(e^{-At}\psi(x_0) - x_0)$$

is

$$x(t) = e^{At}\left(x_0 + W(0,t)W^{-1}(0,T)(e^{-AT}\psi(x_0) - x_0)\right)$$

and hence this control steers x_0 to $\psi(x_0)$ at time T. We will use the trajectories defined in this way to define an equivalent closed loop control. The first step involves solving this equation for x_0. Some manipulation of the equation for $x(t)$ gives the implicit equation

$$x_0 = e^{-At}x(t) - W(0,t)W^{-1}(0,T)(e^{-AT}\psi(x_0) - x_0)$$

Because $W(0,T) \geq W(0,t) \geq 0$ for $0 \leq t \leq T$ we see that $||W(0,T)W^{-1}(0,T)|| \leq 1$. Thus under the given hypothesis the contraction mapping theorem shows that for all t between 0 and T this equation can be solved to get $x_0 = \chi(t,x(t))$. The differentiability of χ follows from the assumed differentiability of ψ and the contraction condition. The control

$$u(t,x) = B^Te^{A^Tt}W^{-1}(0,T)(e^{-At}\psi(\chi(t,x(t)) - \chi(t,x(t)))$$

then steers an arbitrary x_0 to $\psi(x_0)$ at time T.

This proof is analogous to the Picard proof for the existence of solutions for ordinary differential equations in that it proceeds through a sequence of small steps involving a contraction to get a final result that does not involve a contraction.

This lemma has several implications. For example, if ψ is the identity then $e^{-AT}\psi(x) - x$ is a contraction for small values of T and so we can map ρ into itself after an arbitrarily small time. This means that it is possible to maintain ρ near any achievable value of ρ for an arbitrarily long period of time. Similarly, if a diffeomorphism ϕ can be factored as $\phi = \psi_1(\psi_2(...\psi_k(\cdot)...))$ with the ψ_i such that $\psi_i(x) - x$ defines a contraction then we can also reach the diffeomorphism ϕ using a piecewise smooth u. Putting these ideas to gether we have the following result.

Theorem 1: Given $T > 0$, a diffeomorphism $\phi : \mathbb{R}^n \to \mathbb{R}^n$ such that ϕ is the composition of differentiable contractions, and given a controllable linear system

$\dot{x} = Ax + Bu$ on $\mathbb{R}^n$, there exists a piecewise smooth control $u(t,x)$ such that the solution of the Liouville equation

$$\frac{\partial \rho(t,x)}{\partial t} = -\langle \frac{\partial}{\partial x}, (Ax+Bu)\,\rho(t,x)\rangle \;\; ; \;\; \rho(0,x) = \rho_0(x)$$

takes on the value $\rho_0(\phi^{-1}(x))/\det \mathrm{J}_\phi(\mathrm{x})$ at time T.

Remark 1: Linearity of the evolution equation does not play a critical role here. What is important is that that the there should be a solution to the classical controllability problem that depends smoothly on the initial and final states and that the initial condition for the individual trajectories should be uniquely and smoothly determined by the current value of $x(t)$.

Remark 2: It would be of control theoretic interest to understand the controllability of the Liouville equation under the further restriction that u is not allowed to depend on x in its entirety but only allowed to depend on some noninvertible function of x. This situation will be touched on briefly in section 5. It corresponds to what is called *output feedback*, as opposed to *state feedback*.

4. Linear Systems: Controlling the Mean and Variance

In the previous section we showed that the Liouville equation associated with a controllable linear systems is, itself, controllable in a strong sense using feedback control. The control given, however, depends on an arbitrary diffeomorphism reflecting the complexity of the task. In some circumstances it might not be necessary to shape the entire density but only a few moments. We now show how one can get much more explicit results by restricting attention to the control of the first two moments of the solution of the Liouville equation. More specifically, we show that we can control the mean and variance of a linear system using control laws that are affine in x.

Again we consider a controllable system $\dot{x} = Ax+Bu$ together with its Liouville equation and observe that the first two moments can be characterized in terms of

$$\bar{x} = \int_S \rho(t,x)x dx$$

$$\Sigma = \int_S \rho(t,x)(x-\bar{x})(x-\bar{x})^T dx = \int_S \rho(t,x)xx^T dx - \bar{x}\bar{x}^T$$

Making use of the definitions and the equation of evolution with u replaced by $u(t) = K(t)x + v(t)$ we see that

$$\frac{d\bar{x}}{dt} = (A+BK)\,\bar{x} + Bv(t)$$

and

$$\dot{\Sigma} = (A+BK)\,\Sigma + \Sigma\,(A+BK)^T$$

We want to investigate circumstances under which these equations are controllable in the sense that $\bar{x}$ and Σ can be steered to any desired final state consistent with the conditions required by the definitions, $\Sigma(T) = \Sigma^T(T) > 0$. The controllability of the $\bar{x}$ equation follows from the assumption. Now regarding K as a time varying control, the controllability of

$$\dot{\Phi} = (A+BK)\Phi$$

on the space of nonsingular matrices with positive determinant would imply the controllability of the Σ equation on the space of positive definite matrices. This follows from the fact that if the solution of the equation $\dot{\Phi} = F\Phi$ with $\Phi(0) = I$ is $\Phi(T)$ then the solution of

$$\dot{\Sigma} = F\Sigma + \Sigma F^T$$

at time T is $\Phi(T)\Sigma(0)\Phi^T(T)$ and the fact that if Σ_1 and Σ_2 are both positive definite matrices it is possible to find a nonsingular matrix Φ with positive determinant such that $\Phi\Sigma_1\Phi^T = \Sigma_2$.

With this in mind we consider the controllability of $\dot{\Phi} = (A + BK)\Phi$.

Lemma 2: Suppose that (Ω, B) with Ω skew-symmetric is a controllable pair with $e^{\Omega t}$ periodic of period T. Then if $||M|| < 1$ and $\det M > 0$ there exists $K(\cdot)$, a m by n matrix defined on $[0, T]$ such that the solution of the matrix equation

$$\dot{X}(t) = (\Omega + BK(t))X(t) \ \ ; \ \ X(0) = I$$

satisfies $X(T) = M$.

Proof: Recall from the previous section the definition of $W(0, t)$ and the inequality valid for $0 < t \leq T$, $W(0, T) \geq W(0, t) > 0$. The linear matrix equation

$$\dot{X}(t) = \Omega X(t) + BB^T e^{\Omega t} W^{-1}(0, T)(M - I) \ \ ; \ \ X(0) = I$$

has the solution

$$X(t) = e^{\Omega t}\left(I + W(0, t)W^{-1}(0, T)(M - I)\right)$$

and so $X(T) = M$. Consider $\det X(t) = \det(I + W(0, t)W^{-1}(0, T)(M - I))$. On the interval $[0, T]$ we have $W(0, t) \leq W(0, T)$ and so if $||M - I|| < 1$ we have $||W(0, t)W^{-1}(0, T)(M - I)|| < 1$ and on this interval $X(t)$ is nonsingular. Take K to be

$$K(t) = B^T e^{\Omega t} X^{-1}(t)$$

The following lemma asserts that we can get essentially the same conclusion with significantly weaker assumptions.

Lemma 3: Let T be positive and let (A, B) be a controllable pair. Then if $||M||$ has a positive determinant there exists $K(\cdot)$ defined on $[0, T]$ such that the solution of the matrix equation

$$\dot{X}(t) = (A + BK(t))X(t) \ \ ; \ \ X(0) = I$$

satisfies $X(T) = M$.

Proof: Rather than looking at this problem as a control problem involving right invariant vector fields on a Lie group as in [2], we continue along the lines of the previous lemma. Let $T > 0$ be given. It is well known in control theory that under the given hypothesis it is possible to find a real, constant matrix K_0 such that the eigenvalues of $A + BK_0$ equal any given subset of n complex numbers subject only to the constraint that the complex eigenvalues occur in conjugate pairs. We write K as $K_0 + K_1$ with K_0 such that the eigenvalues of $A + BK_0$ are purely imaginary, unrepeated and a multiple of $2\pi/T$. Under this assumption we can find P such that $P(A + K_0)P^{-1} = \Omega$ with (Ω, B) satisfying the assumptions of the lemma. Thus $Y = PXP^{-1}$ satisfies

$$\dot{Y}(t) = (\Omega + PBK(t)P^{-1})Y(t) \ \ ; \ \ Y(0) = I$$

This result, together with the previous discussion allows us to claim the following result on the controllability of the mean and variance associated with a linear system.

Theorem 2: Let T be positive and let (A, B) be a controllable pair. Then for any given m and $\Sigma_1 > 0$ there exists $K(\cdot)$ and v defined on $[0, T]$ such that the solution of

$$\frac{d}{dt}\bar{x}(t) = (A + BK(t))\bar{x}(t) + Bv \;\; ; \;\; x(0) = x_0$$

$$\dot{\Sigma}(t) = (A + BK(t))\Sigma + \Sigma(A + BK(t))^T \;\; ; \;\; \Sigma(0) > 0$$

satisfies $x(T) = m$, $\Sigma(T) = \Sigma_1$.

Remark 2: It is of interest to explore the extent to which the affine control law can be restricted to laws of the form $u = KCx$ with C constant but not invertible. As mentioned above, in the control theory literature this would be called the output feedback case to distinguish it from the state feedback case treated above. In this case we have

$$\frac{d\bar{x}}{dt} = (A + BKC)\,\bar{x} + Bv(t)$$

and

$$\dot{\Sigma} = (A + BKC)\,\Sigma + \Sigma\,(A + BKC)^T$$

and again the controllability of the $\bar{x}$ equation follows from the assumption. The controllability of

$$\dot{\Phi} = (A + BKC)\Phi$$

on the space of nonsingular matrices with positive determinant would imply the controllability of the Σ equation on the space of positive definite matrices but now the previous argument does not apply and it seems we must treat this as a controllability problem on Lie groups, as in [3], and be content with less complete results.

5. Optimal control

Having established the existence of controls that steer the mean and variance appropriately, we now turn to a question involving optimal control. In this section we consider the more general output feedback situation but limit ourselves to the so called normal problems of the maximum principle.

Theorem 3: Given $T > 0$, a controllable pair (A, B) and n by n matrices X_0 and X_f with $\det X_f X_0 > 0$, suppose that the control $K(t)$ steers the system

$$\dot{X} = (A + BKC)X \;\; ; \;\; X(0) = X_0$$

to X_f in T units of time while minimizing

$$\eta = \int_0^T \frac{1}{2}||K^T K||^2 dt$$

Then if the problem is normal

$$K(t) = -B^T M(t) C^T$$

with M satisfying the quadratic, isospectral equation

$$\dot{M} = [M, (A - BB^T M C^T C)^T]$$

Proof: Assume an optimal control exists and the problem is normal. According to the maximum principle there is a matrix P satisfying

$$\dot{P} = -(A + BKC)^T P$$

such that the optimal value of K minimizes $\mathrm{tr}(P^T(A+BKC)X+\frac{1}{2}K^TK)$. Thus the optimal K is given by $K=-B^TPX^TC^T$. A short calculation shows that $M=PX^T$ satisfies the given equation. Any matrix flow of the form $\dot{M}=[M,\psi(M)]$ evolves in such a way that the spectrum of M does not change.

In view of the fact that transpose inverse of X satisfies the same equation as P, there exists a constant matrix R such that $P=(X^{-1})^TR^T$ to get $M^T=XR^{-1}X^{-1}$ which implies the following

Corollary: Under the hypothesis of the theorem, there exists a constant matrix R such that the optimal trajectories satisfy

$$\dot{X}=(A-BB^T(XR^{-1}X^{-1})^TC^TC)X$$

Remark 3: This result can be recast as a result about the variance equation in the following way. Given $T>0$, a controllable pair (A,B) and n by n symmetric, positive definite matrices Σ_0 and Σ_f, suppose that the control $K(t)$ steers the system

$$\dot{\Sigma}=(A+BKC)\Sigma+\Sigma(A+BKC)^T \;\;;\;\; \Sigma(0)=\Sigma_0$$

to Σ_f in T units of time while minimizing

$$\eta=\int_0^T ||K^TK||^2dt$$

Then for some value of $M(0)$

$$K(t)=-B^TM(t)C^T$$

with

$$\dot{M}=[M,(A-BB^TMC^TC)^T]$$

We point out that if one applies these ideas directly to the equations for the mean and variance the result changes only slightly. This is the promised generalization of Example 2 of Section 2.

Theorem 4: Given $T>0$, a controllable linear system $\dot{x}=Ax+Bu$, two symmetric, positive definite matrices $\Sigma(0)$ and Σ_f and vectors $\bar{x}(0),\bar{x}_f$, if the control law (v,K) steers the system

$$\frac{d}{dt}\bar{x}=(A+BKC)\bar{x}+Bv \;\;;\;\; \bar{x}(0)=\bar{x}_0$$

$$\dot{\Sigma}=(A+BKC)\Sigma+\Sigma(A+BKC)^T \;\;;\;\; \Sigma(0)=\Sigma_0$$

to $(\bar{x}_f,\Sigma_f)$ in T units of time while minimizing

$$\eta=\int_0^T \frac{1}{2}v^Tv+\mathrm{tr(K^TK)dt}$$

then

$$K=-B^T(M+qx^T)C^T \;\;;\;\; v=-B^Tq(t)$$

with q and M satisfying

$$\dot{q}=-(A+BKC)^Tq$$

$$\dot{M}=[M,(A-BB^TMC^TC)^T]$$

6. Generating Elements of Diff $^{(1)}(X)$ with $\dot{x} = f(x,u)$

Theorem one describes a strong controllability property of the Liouville equation associated with a controllable linear system. Its proof uses the fact that we have a convenient explicit expression for a control that executes a given transfer. The development of results applicable to more general controllable systems of the form

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

requires different techniques because no such explicit expression is generally available. However, in this more general setting it is useful to observe that the reachable set for solutions of the Liouville equation can sometimes be better investigated by observing that it is invariant under the transformation $(f(x), G(x)) \mapsto (f(x) + G(x)p(x), G(x)Q(x))$ with $p(x)$ a vector of appropriate smoothness and Q smooth and invertible. This is often described by saying that the reachable set is invariant under the action of the fedback group. (See [4] for elaborations on this theme.) With this in mind we recast the situation described in theorem 1 in terms of vector fields and Lie algebras.

Theorem 5: Let (A, B) be a controllable pair in dimension n. Given a $\mathcal{C}^1$ vector field on $\mathbb{R}^n$,

$$F = f_1(x)\frac{\partial}{\partial x_1} + f_2(x)\frac{\partial}{\partial x_1} + \cdots + f_n(x)\frac{\partial}{\partial x_n}$$

there exists a finite collection of $\mathcal{C}^1$ functions $\phi_1, \phi_2, ..., \phi_m$ such that F belongs to the Lie algebra generated by

$$S = \left\{ \sum_i a_{ij}x_j \frac{\partial}{\partial x_i}, \sum_i b_{ij}\phi_k \frac{\partial}{\partial x_i} \right\}_{k=1,2,\ldots,m}$$

Proof: Let B have columns $b_1, b_2, ...b_m$. It is convenient to represent the vector fields as vectors in $\mathbb{R}^n$, i.e., as Ax, b_i etc. Consider the Lie bracket of Ax and $b_i\phi_j$

$$[Ax, b_i\phi_j] = Ab_i\phi_j - b_i\langle d\phi_j, Ax\rangle$$

and so if we include along with each ϕ_i the function $\langle d\phi_i, Ax\rangle$ we see that we have $Ab_i\phi_j$ in S. Likewise, the bracket

$$[Ax, Ab_i\phi_j] = A^2b_i\phi_j - Ab_i\langle d\phi, Ax\rangle$$

so $A^2b\phi_j$ belongs to the Lie algebra, etc. In this way we see that $A^kb\phi_j$ belongs to the Lie algebra for each k. In view of the Cayley-Hamiltonian theorem and the controllability assumption, as k ranges from 0 to $n-1$ the set $A^kb_1, A^kb_2, ...$ contains a basis for $\mathbb{R}^n$ so we have proven the theorem.

This provides another point of view on situation treated in theorem one pointing out that controllable linear systems have a very strong vector field generating property.

References

[1] R. W. Brockett, *Minimum Attention Control*, Proceedings of the 1997 Conference on Decision and Control, IEEE, Piscataway, NJ, 1997, pp. 2628–2632.

[2] Jurgen Moser, *On the Volume Elements of a Manifold.* Trans. American Math. Soc., **120** (1965), 286–294.

[3] R. W. Brockett, *System Theory on Group Manifolds and Coset Spaces.* SIAM Journal on Control, **10** (1972), 265–284.

[4] R. W. Brockett, *Feedback Invariants for Nonlinear Systems.* Proceedings of the 1978 IFAC Congress, Helsinki, Finland, Pergamon Press, Oxford, 1978.

Engineering and Applied Science, Harvard University
E-mail address: brockett@deas.harvard.edu

ISSN 0081-5438, Proceedings of the Steklov Institute of Mathematics, 2010, Vol. 268, pp. 49–57. © Pleiades Publishing, Ltd., 2010.

On the Control of a Flock by a Leader

R. W. Brockett [a]

Received June 2009

Abstract—The collective and individual behaviors of the elements of a set of identical units are investigated from the point of view of their response to a coordination signal broadcast by a leader. In terms of the model used here, we show that for such systems nonlinearity plays a critical role. A new method for establishing controllability of nonlinear, replicated systems is given. Similarly, stabilization also depends on high order nonlinear effects. For the purpose of stabilization, a distinction between odd order nonlinear terms and even order nonlinear terms plays a role.

DOI: 10.1134/S0081543810010050

1. INTRODUCTION

Problems involving the coordination of the movement of flocks have been the subject of numerous recent investigations, motivated both by a desire to better understand naturally occurring ensembles and by the possible uses for man made systems of this type (see [1]). Our goal here is to establish two important properties relating to the responsiveness of a set of identical dynamical systems when a leader attempts to influence their behavior through "broadcast" signals. We introduce a dynamical model for the motion of a set of masses moving together in a cluster while trying to maintain a regular spacing between themselves, such as might describe formation flight of aircraft or birds in $\mathbb{R}^3$ or a set of ships in $\mathbb{R}^2$. Our models are elementary but yet detailed enough to reveal some rather subtle effects. In particular, the analysis of the possibilities for stabilization can be delicate, requiring the consideration of nonlinear effects. This question can be considered as a discrete form of the problem of controlling the corresponding Liouville equation as in [2] but here we take a different approach.

Some of the difficulties which arise in controlling such systems arise exactly because of the assumption of identical units. For example, observe that if we have an oscillator in $\mathbb{R}^k$, described by $\ddot{x} + Qx = 0$, $Q = Q^{\mathrm{T}}$, then a rank one damping term $-bc^{\mathrm{T}}\dot{x}$ can stabilize the system if the eigenvalues of Q are not repeated but not otherwise. In fact, if the eigenvalues of Q are all the same, then the addition of the derivative term shifts just one eigenvalue off the imaginary axis. This suggests that the control of identical systems with broadcast signals might be delicate.

Although the literature contains remarks about the lack of importance of a leader in biological situations, the arguments given against assuming an important role for a leader are based on observational studies rather than analysis and for this reason it seems worthwhile to make a mathematical study of the situation. As part of our treatment we will make use of an artificial periodic potential, which serves to provide a degree of global coordination. We use the fact that the location of the minima of such a periodic potential defines a virtual lattice. One can think of this potential as being generated either by nearest neighbor interactions or by some longer range "collective" effect, the exact nature of which we are content to leave vague here. Like the use of lattice potentials in solid state physics, where they arise, for example, as an electrostatic field which influences the

[a] School of Engineering and Applied Sciences, Harvard University, 29 Oxford Street, Cambridge, MA 02138, USA.
E-mail address: brockett@seas.harvard.edu

movement of electrons and is created by charged ions in a crystal, only some of the lattice sites need to be populated. The widely studied graph rigidity results [3–5] can be seen as providing limitations on how connected the occupied sites must be.

2. THE VIRTUAL LATTICE MODEL

Consider a potential function $\phi\colon \mathbb{R}^p \to \mathbb{R}$ which is periodic in the sense that there exists p linearly independent vectors $\xi_1, \xi_2, \dots, \xi_p$ such that $\phi(x+\xi_j) = \phi(x)$ for each of the ξ_j. Let Γ be the corresponding group of translations acting on $\mathbb{R}^n$ so that $\phi(\Gamma_i x) = \phi(x)$ for each $\Gamma_i \in \Gamma$. A unit mass moving in $\mathbb{R}^p$ under the influence of such a potential satisfies the second order equation

$$\ddot{x} + \nabla\phi(x) = 0,$$

and if the initial kinetic energy is smaller than the difference between the maximum value of the potential and the initial potential energy, the mass will execute a bounded motion remaining near its initial position. Moreover, if there are k such particles, each starting near a different local minimum and having suitably small kinetic energy, then they will oscillate independently, remaining near their starting point. Furthermore, if the potential is moving rigidly, which is to say if ϕ is replaced by

$$\phi(x,t) = \phi(x - d(t)),$$

then the change of variables $w = x - d(t)$ results in a description of the relative motion of the form

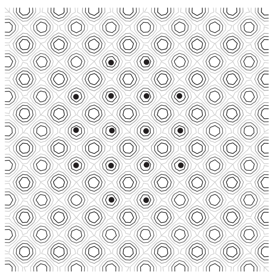
$$\ddot{w} + \nabla\phi(w) = \ddot{d}.$$

Again, we may have multiple unit masses occupying positions near distinct local minima. The lattice serves as a template creating separation between the masses while assuring a degree of coordination. We refer to it as being virtual because it is pictured as being induced not by physical law but rather by discipline or animal instinct.

The questions to be discussed here relate to the extent to which the positions of the individual masses can be influenced by a single "broadcast" signal. In particular, whether the motions can be simultaneously stabilized by a single feedback signal. We are especially interested in broadcast signals which are symmetric functions of the positions and velocities of the individual masses.

The assumptions of identical units and a periodic potential mean that the composite system has a number of symmetries. This can be expressed in terms of the overall system by saying that there is a finite subgroup of the Euclidean group with elements $P_1, P_2, \dots, P_m$ such that if the overall description is $\dot{x} = f(x) + ug(x)$, then

$$P_i f(P_i^{-1}x) = f(x), \qquad P_i g(P_i^{-1}x) = g(x), \qquad f(x+\xi_i) = f(x).$$



Contours of the lattice defining the potential $\phi(x) = \sin x_1 + \sin x_2$ and an ensemble of 16 units near equilibrium.

What we will see is that for linear systems obtained by linearizing about any one of the possible equilibria, this type of symmetry precludes controllability and feedback stabilization, making nonlinear models essential.

Remark. If $P_1, P_2, \dots, P_m$ are the elements of a nontrivial finite group of matrices and if a linear system of the form $\dot{p} = Ap + bu$ is symmetric in the sense that $P_iA = AP_i$, $P_ib = b$ for all $i = 1, 2, \dots, m$, then the system is not controllable. To see this observe that $P_i(b, Ab, \dots, A^{n-1}b) = P_j(b, Ab, \dots, A^{n-1}b)$ and so $(P_i - P_j)(b, Ab, \dots, A^{n-1}b) = 0$ and $P_i - P_j$ is nonzero if the group is nontrivial. Thus linear systems with this type of symmetry are never controllable in broadcast mode. This type of analysis is treated in more detail in [6].

In an expanded theory it would be natural to include terms in the potential that are not periodic, e.g.,

$$\phi(x, t) = \phi(x - d(t)) + v(t).$$

Such a term could be used to reshape the configuration of the units so as to maneuver around obstacles, etc., and could be used to shift the units from one lattice site to another.

3. CONTROLLABILITY

When considering the control of identical systems via a single "broadcast" control, it is convenient to have a notation for a system consisting of a number of independent, identical systems. Given a positive integer k, we associate with the control system $\dot{x} = f(x) + g(x)u$ in $\mathbb{R}^n$ a system evolving in $\mathbb{R}^{kn}$ described by

$$\dot{p}_x = L_f^k + L_g^k u$$

with the definitions

$$\dot{p}_x = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_k \end{bmatrix}, \qquad L_f^k = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_k) \end{bmatrix}, \qquad L_g^k = \begin{bmatrix} g(x_1) \\ g(x_2) \\ \vdots \\ g(x_k) \end{bmatrix}.$$

It follows immediately that

$$[L_f^k, L_g^k] = L_{[f,g]}$$

and so any Lie algebraic approach to controllability questions will involve the study of the Lie algebra generated by f and g, which we denote as $\{f, g\}_{\mathrm{LA}}$. Of course, the controllability of the individual systems $\dot{x}_i = f(x) + g(x_i)u$ is necessary for the composite system to be controllable but it is far from sufficient.

The dimension of the distribution in $\mathbb{R}^{nk}$ generated by the Lie algebra of vector fields L_f^k and L_g^k can be expected to decrease if, for example, $x_i = x_j$. Aside from this kind of degeneracy which occurs on lower dimensional subsets, we can expect to determine conditions on f and g which result in the distribution generated by L_f^k and L_g^k being of dimension nk at "generic" points. We assume, for the rest of this section, that f and g are infinitely differentiable.

Lemma 1. *The dimension of the distribution generated by L_f^k and L_g^k does not exceed the number of $\mathbb{R}$-linearly independent vector fields in $\{f, g\}_{\mathrm{LA}}$.*

Proof. Suppose that u is piecewise constant. In a neighborhood of any initial value of x we can represent the solution of $\dot{x} = f(x) + ug(x)$ as the composition of exponentials of the Lie algebra generated by the vector fields f and g. If we let $\alpha_1, \alpha_2, \dots, \alpha_r$ be a basis, then

$$x(t) = \exp(\exp(\dots(\exp(\alpha_r u_r)\alpha_{r-1}u_{r-1})\dots)\alpha_1 u_1)x(0)$$

and for p we have

$$p(t) = \exp(\exp(\ldots(\exp(L^k_{\alpha_r}u_r)L^k_{\alpha_{r-1}}\alpha_{r-1}u_{r-1})\ldots)L^k_{\alpha_1}u_1)x(0).$$

This shows that the solutions generated by piecewise constant controls are contained in a manifold of the required dimension. However, the piecewise constant controls generate solutions that are dense in the space of all solutions and therefore the claim is established.

Notice that in the case of a controllable linear system with $f(x) = Ax$ and $g(x) = b$, with A and b constant, the Lie algebra generated by f and g has a basis consisting of $\{Ax, b, Ab, \ldots, A^{n-1}b\}$ and thus is of dimension $n+1$. The dimension of the distribution generated by L^k_f and L^k_g cannot exceed $n+1$, regardless of k. In the case of a controllable system with $f(x) = Ax$ and $g(x) = Bx$, with A and B constant, the dimension of the Lie algebra $\{Ax, Bx\}_{\mathrm{LA}}$ cannot exceed n^2. Thus $\dot p = L^k_f + uL^k_g$ has less restricted controllability properties as compared with the standard linear situation discussed above; however, the system cannot be controllable unless $k \le \frac{1}{n}\dim\{A, B\}_{\mathrm{LA}}$ because the solution of $\frac{d}{dt}\Phi = \bigl(A + \sum u_i B_i\bigr)\Phi$ is a $\dim\{A, B\}_{\mathrm{LA}}$ parameter family.

Two theorems below give conditions under which the generic dimension of the reachable set for $\dot p = L_f + uL_g$ is, in fact, nk. The results are based on the following elementary lemma giving a property of linear systems.

Lemma 2. *For $i = 1, 2, \ldots, k$, let A_i and b_i be $n \times n$ matrices and n-vectors, respectively. Assume that for each i, b_i is a cyclic vector for A_i. Let $\Lambda(A_i)$ denote the set of eigenvalues of A_i. Then the kn-dimensional linear system*

$$\begin{bmatrix} \dot x_1 \\ \dot x_2 \\ \vdots \\ \dot x_k \end{bmatrix} = \begin{bmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & A_k \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} u$$

is controllable if and only if the intersection of $\Lambda(A_i)$ and $\Lambda(A_j)$ is empty for all $i \ne j$.

Proof. A necessary and sufficient condition for a single input linear system $\dot x = Ax + bu$ to be controllable is that the components of the vector $e^{At}b$ be linearly independent over any interval of positive length. These components consist of linear combinations of $t^{r_s}e^{\lambda_s t}$ with λ_i being the eigenvalues of $A_1, A_2, \ldots, A_k$ and r_s taking on values from zero to the multiplicity of λ_s. But as is well known, such functions are independent on any interval of positive length provided that the λ_s are pairwise distinct.

We now use this lemma to establish an easily checkable test for determining when L^k_f and L^k_g generate a distribution of dimension nk at generic points. Notice that because of our continuity assumptions, the dimension of the distribution associated with a set of vector fields can be thought of as the rank of a matrix whose entries are continuous functions of x and hence the rank is an upper semi-continuous function of the point at which the vector fields are evaluated.

Theorem 1. *Let N be an open set in $\mathbb{R}^n$ and let f and g be $\mathcal{C}^\infty$ vector fields defined on N. Let $\gamma_0 \in N$ be such that for $A = \partial f/\partial x|_{\gamma_0}$ and $b = g(\gamma_0)$ the pair A, b is controllable. If there are k points $\gamma_1, \gamma_2, \ldots, \gamma_k$ in N such that for all $i \ne j$ the eigenvalues of $A_i = \partial f/\partial x|_{\gamma_i}$ and $A_j = \partial f/\partial x|_{\gamma_j}$ are disjoint in the sense of Lemma 1, then at the point $p_0 = [\gamma_1, \gamma_2, \ldots, \gamma_k] \in N \times N \times \ldots \times N$ the dimension of the distribution generated by L^k_f and L^k_g is nk.*

Proof. To the first order in t and u the solution of $\dot p = L^k_f(p) + L^k_g(p)u$ is given by

$$p(t) = e^{A^k_f t}p(0) + \int_0^t e^{A^k_f(t-\sigma)}g(p(0))u(\sigma)\,d\sigma + L^k_f(p_0)t$$

where

$$A_f^k = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_k \end{bmatrix}$$

with

$$A_i = \left.\frac{\partial f}{\partial x}\right|_{\gamma_i}.$$

By assumption this first order approximation is controllable and hence the distribution at the point p_0 is of dimension nk.

The task of finding nk linearly independent vector fields is sometimes greatly simplified by the following theorem in which the second derivative of f evaluated at a single point is key.

Theorem 2. *Let N be an open set in $\mathbb{R}^n$ and let f and g be $\mathcal{C}^\infty$ vector fields defined on N. Let $\gamma_0 \in N$ be such that $f(\gamma_0) = 0$. Assume that for $A = \partial f/\partial x|_{\gamma_0}$ and $b = g(\gamma_0)$ the pair A, b is controllable and that the eigenvalues of A are nonzero and unrepeated. Let ξ_i be a spanning set of right eigenvectors for A_0 and let χ_i be the corresponding left eigenvectors; $A_0\xi_i = \lambda_i\xi_i$ and $\chi_i^{\mathrm{T}} A_0 = \chi_i^{\mathrm{T}}\lambda_i$. For $\delta \in \mathbb{R}^n$, define the matrix B_δ with the ij-th entry*

$$(B_\delta)_{ij} = \sum_{k=1}^{n} \left.\frac{\partial^2 f_i}{\partial x_j \partial x_k}\right|_{\gamma_0} \delta_k.$$

Assume that for all δ in N we have

$$\chi_i^{\mathrm{T}} B_\delta \xi_i \neq 0, \qquad i = 1, 2, \dots, n.$$

Then there are points in $N \times N \times \dots \times N$ such that the distribution generated by L_f and L_g is nk-dimensional.

Proof. Fix a value of δ. From the hypothesis we have

$$\left.\frac{\partial f}{\partial x}\right|_{\gamma_0+\delta} = A + B_\delta + \dots .$$

Standard eigenvalue perturbation analysis leads to the first order approximation for the eigenvalues of $A + B_{\epsilon\delta}$ as $\lambda_i + \epsilon\nu_i$ with

$$\nu_i \chi_i^{\mathrm{T}} \xi_i = \chi_i^{\mathrm{T}} B_\delta \xi_i.$$

Note that the hypothesis implies that $\chi_i^{\mathrm{T}}\xi_i \neq 0$. Let $\epsilon_1, \epsilon_2, \dots, \epsilon_k$ be scalars and consider the matrix

$$\mathcal{A} = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_k \end{bmatrix}$$

where now $A_j = A + \epsilon_j B_\delta$. Label the ith eigenvalue of the jth block λ_{ij} so that to the first order in ϵ_j we have

$$\lambda_{ij} = \lambda_i + \epsilon_j \frac{\chi_i^{\mathrm{T}} B_\delta \xi_i}{\chi_i^{\mathrm{T}} \xi_i}.$$

Using the assumption that $\chi_i^{\mathrm{T}} B_\delta \xi_i \neq 0$, we see that there is a set of constants $[\epsilon_1, \epsilon_2, \dots, \epsilon_k]$ such that at $p = [\gamma_0 + \epsilon_1\delta, \gamma_0 + \epsilon_2\delta, \dots, \gamma_0 + \epsilon_k\delta]$ the eigenvalues of $\mathcal{A}$ are distinct and thus the distribution generated by L_f^k and L_g^k is nk-dimensional.

The following example illustrates how effective these ideas can be.

Example. Consider f and g defined by

$$\begin{bmatrix} \dot{x} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} v \\ -f(x) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

with $f(0) = 0$, $f'(0) = 1$ and $f''(0) = a \neq 0$. Let γ_0 of Theorem 2 be 0 so that

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad B_\delta = \begin{bmatrix} 0 & 0 \\ a\delta & 0 \end{bmatrix}.$$

Then the right and left eigenvectors of A are given by

$$\xi = \begin{bmatrix} 1 \\ \pm\sqrt{-1} \end{bmatrix}, \qquad \chi = \begin{bmatrix} 1 \\ \pm\sqrt{-1} \end{bmatrix}.$$

The quantities

$$\chi_i^{\mathrm{T}} B_\delta \xi_i = \pm 2a\delta i.$$

Then for all $k = 1, 2, \dots$ the corresponding $2k$-dimensional system $\dot{p} = L_f^k(p) + uL_g^k(p)$ is controllable if $f''(0) \neq 0$.

Remark. In Theorem 2 it is possible to relax the assumptions on the eigenvalues of A at the expense of a more complicated condition on B.

4. STABILITY

Here we turn to the question of stabilization of systems of the form discussed above. The fact that controllability is lost along subspaces of the form $x_i \equiv x_j$, $x_i \equiv x_j \equiv x_k$, etc., should not be viewed as a deterrent to stabilization because all the x_i are to be driven to the same point.

Lemma 3. *Let $f\colon \mathbb{R}^1 \to \mathbb{R}^1$ be $f(x) = ax + bx^2$ with $a < 0$ and $b \neq 0$. Consider the family of k identical second order differential equations*

$$\ddot{x}_1 = f(x_1) + u,$$
$$\ddot{x}_2 = f(x_2) + u,$$
$$\dots$$
$$\ddot{x}_k = f(x_k) + u.$$

If

$$u = -(\dot{x}_1 + \dot{x}_2 + \ldots + \dot{x}_k),$$

then the null solution is locally asymptotically stable.

Proof. Consider the Liapunov function

$$v(x) = \sum_{j=1}^{k} \left(\frac{1}{2}\dot{x}_j^2 - \frac{a}{2}x_j^2 - \frac{b}{3}x_j^3 \right),$$

which is positive definite in a neighborhood of the origin. An easy calculation shows that

$$\dot{v} = -\left(\sum_{j=1}^{k} \dot{x}_j\right)^2.$$

By the Krasovskii–LaSalle's theorem [7, 8], it follows that the null solution is asymptotically stable provided that the derivative of v is not identically zero along any nonzero trajectory. We now show that if it is identically zero, then each of the x_j is identically zero. If $\dot{v}$ is identically zero, the equations for the x_j decouple; $\ddot{x}_j = ax_j + bx_j^2$. Clearly each x_j is infinitely differentiable. Letting x denote a typical x_j, the first few higher derivatives are

$$\begin{aligned}
\frac{d^3x}{dt^3} &= \dot{x}(a + 2bx),\\
\frac{d^4x}{dt^4} &= a^2x + 3abx^2 + 2b^2x^3 + 2b\dot{x}^2,\\
\frac{d^5x}{dt^5} &= \dot{x}(a^2 + 10abx + 10b^2x^2).
\end{aligned}$$

We can express the ith derivative of x as a finite sum

$$\frac{d^ix}{dt^i} = \sum m^i_{\alpha,\beta} x^\alpha \dot{x}^\beta, \qquad \alpha > 0, \quad \beta \ge 0.$$

Introduce the operator

$$D = \left(\dot{x}\frac{\partial}{\partial x} + f\frac{\partial}{\partial \dot{x}}\right),$$

which maps the vector space consisting of finite sums of polynomials of this type into itself. This operator characterizes the relationship between the coefficients in the expression for d^ix/dt^i and the coefficients in the expression for $d^{i+1}x/dt^{i+1}$. It acts on the generic term in accordance with

$$\left(\dot{x}\frac{\partial}{\partial x} + f\frac{\partial}{\partial \dot{x}}\right) m^i_{\alpha,\beta} x^\alpha \dot{x}^\beta = m^i_{\alpha,\beta}\alpha x^{\alpha-1}\dot{x}^{\beta+1} + m^i_{\alpha,\beta}\big(ax^{\alpha+1} + bx^{\alpha+2}\big)\dot{x}^{\beta-1}.$$

Define $\mathrm{degree}_\nu(x^\alpha\dot{x}^\beta)$ to be $\alpha + 3\beta/2$ and notice that for $b \neq 0$, D acts to raise degree_ν by $1/2$,

$$\mathrm{degree}_\nu(D(x^\alpha\dot{x}^\beta)) = \mathrm{degree}_\nu(x^\alpha\dot{x}^\beta) + 1/2.$$

Reasoning inductively from the expression for the second derivative, $\ddot{x} = ax + bx^2$, we see that

(i) the terms in the expression for d^q/dt^q have degree_ν less than or equal to $(q+2)/2$ and

(ii) the odd time derivatives of x depend only on odd powers of $\dot{x}$ and the even time derivatives of x depend only on even powers of $\dot{x}$.

Putting these remarks together, we see that for q odd the highest power of x multiplying $\dot{x}$ in the expression for d^q/dt^q is $x^{(q-1)/2}$ and thus

$$\frac{d^q}{dt^q}x = c\dot{x}(bx)^{(q-1)/2} + \dot{x}\chi(x) + \psi(x,\dot{x})$$

where $\psi(x,\dot{x})$ is of degree 3 or higher in $\dot{x}$ and $\chi(x)$ is of degree $(q-3)/2$ or lower in x.

We now organize these results to show that if $\sum \dot{x}_i \equiv 0$, then the individual $\dot{x}_i$ must be zero. If $\sum \dot{x}_i \equiv 0$, then all its higher derivatives are zero as well. It is convenient to consider only the odd derivatives. Using the above calculations, we can organize the first k such equations as

$$\begin{bmatrix} \sum \frac{d^1 x_i}{dt^1} \\ \sum \frac{d^3 x_i}{dt^3} \\ \vdots \\ \sum \frac{d^{2k-1} x_i}{dt^{2k-1}} \end{bmatrix} = \begin{bmatrix} m_1(x_1) & m_1(x_2) & \dots & m_1(x_k) \\ m_2(x_1) & m_2(x_2) & \dots & m_2(x_k) \\ \vdots & \vdots & \ddots & \vdots \\ m_k(x_1) & m_k(x_2) & \dots & m_k(x_k) \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_k \end{bmatrix} + \xi(x_i, \dot{x}_i)$$

with ξ being of order 3 and higher in the variables $\dot{x}_i$. If $\dot{v}$ vanishes identically, then

$$\begin{bmatrix} m_1(x_1) & m_1(x_2) & \dots & m_1(x_k) \\ m_2(x_1) & m_2(x_2) & \dots & m_2(x_k) \\ \vdots & \vdots & \ddots & \vdots \\ m_k(x_1) & m_k(x_2) & \dots & m_k(x_k) \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_k \end{bmatrix} + \xi(x_i, \dot{x}_i) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

We now put to use what we have shown about the m_i. Clearly, $m_1(x_i) = 1$ and $m_2(x_i) = a + 2bx_i$. From the recursions developed above we see that

$$m_q(x_i) = \dot{x}_i(bx_i)^{(q-1)/2} + \dot{x}_i\phi_q(x_i) + \psi_q(x_i, \dot{x}_i)$$

with the polynomial $\phi_q(x_i)$ being of degree $(q-3)/2$ or less. By subtracting a suitable linear combination of the first $q-1$ rows from the qth row, we arrive at an equivalent system in which the matrix multiplying the $\dot{x}_i$ takes the form of a Vandermonde matrix. Its determinant vanishes only if two or more of the solutions are identical, and the inverse function theorem assures us that if no two solutions are equal, then for small enough values of $\dot{x}_i$ the obvious solution $\dot{x}_i \equiv 0$ is the only solution. Now divide the possibilities into two cases. If we assume that there are no such identical solutions, then the system defined above has a unique solution for the $\dot{x}_i$ in a neighborhood of zero and that solution is the solution $\dot{x}_i \equiv 0$. Thus in this case we have established the lemma. On the other hand, if two of the x_i are identical, we can replace the above system by a system with one fewer unknown, resulting in a smaller system with the same structure. This provides an inductive step which, after a finite number of repetitions, results in reducing the situation to the first case. This completes the proof.

This lemma has the following implication.

Theorem 3. *Consider the system*

$$\begin{bmatrix} \dot{x} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} v \\ -f(x) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

with f infinitely differentiable, $f(0) = 0$, $f'(0) > 0$ and $f''(0) \neq 0$. Then for all $k = 1, 2, \dots$ the null solution of the corresponding $2k$-dimensional system $\dot{p} = L^k_{f+ug}$ is locally asymptotically stable for

$$u = -(v_1 + v_2 + \dots + v_k).$$

Proof. Massera's theorem [9] asserts that for the system described in Lemma 3 there is a positive definite Liapunov function with a negative definite derivative. Using this Liapunov function with the system here, we see that in a neighborhood of 0, the contribution of the terms involving powers of the third and higher order in f are dominated by the negative definite term and therefore the system is (locally) asymptotically stable.

ACKNOWLEDGMENTS

This work was supported in part by the US Army Research Office under grant DAAG 55 97 1 0114.

REFERENCES

1. C. W. Reynolds, "Flocks, Herds, and Schools: A Distributed Behavioral Model," Computer Graphics **21** (4), 25–34 (1987) (SIGGRAPH'87 Conf. Proc.).
2. R. W. Brockett, "Optimal Control of the Liouville Equation," in *Proc. Int. Conf. on Complex Geometry and Related Fields*, Ed. by Zhijie Chen et al. (Am. Math. Soc., Providence, RI, 2007), pp. 23–35.
3. T. Eren, P. N. Belhumeur, and A. S. Morse, "Closing Ranks in Vehicle Formations Based on Rigidity," in *Proc. 41st IEEE Conf. on Decision and Control* (IEEE Press, Piscataway, 2002), Vol. 3, pp. 2959–2964.
4. R. Olfati-Saber and R. M. Murray, "Graph Rigidity and Distributed Formation Stabilization of Multi-vehicle Systems," in *Proc. 41st IEEE Conf. on Decision and Control* (IEEE Press, Piscataway, 2002), Vol. 3, pp. 2965–2971.
5. T. Eren, D. K. Goldenberg, W. Whitley, Y. R. Yang, A. S. Morse, B. D. O. Anderson, and P. N. Belhumeur, "Rigidity, Computation, and Randomization in Network Localization," in *Proc. INFOCOM 2004: 23rd Annual Joint Conf. IEEE Computer and Commun. Soc., Hong Kong, China, Mar. 2004* (IEEE Press, Piscataway, 2004), Vol. 4, pp. 2673–2684.
6. R. W. Brockett, "The Global Description of Locally Linear Systems," in *Feedback Control of Linear and Nonlinear Systems*, Ed. by D. Hinrichsen and A. Isidori (Springer, New York, 1982), Lect. Notes Control Inf. Sci. **39**, pp. 1–8.
7. N. N. Krasovskii, *Some Problems of the Theory of Stability of Motion* (Fizmatgiz, Moscow, 1959). Engl. transl.: N. N. Krasovskii, *Stability of Motion. Applications of Lyapunov's Second Method to Differential Systems and Equations with Delay* (Stanford Univ. Press, Stanford, CA, 1963).
8. J. P. LaSalle, "Some Extensions of Liapunov's Second Method," IRE Trans. Circuit Theory **CT-7**, 520–527 (1960).
9. J. L. Massera, "On Liapounoff's Conditions of Stability," Ann. Math., Ser. 2, **50**, 705–721 (1949).

This article was submitted by the author in English