Optimal Control for Nonlinear Processes

E. B. LEE & L. MARKUS

Communicated by J. SERRIN

Introduction

In an optimum control problem we are given a real ordinary differential equation system

$$\frac{dx^i}{dt} = f^i(t; x^1, x^2, \ldots, x^n; u^1, \ldots, u^m) \qquad i = 1, 2, \ldots, n$$

which is a mathematical model of some physical process. The problem of control is to select the real functions $u^j(t)$, $j = 1, 2, \ldots, m$ (control variables) on an interval of time, $t_0 \leqq t \leqq t_1$, such that the solution $x^i(t)$ moves in a prescribed manner on $t_0 \leqq t \leqq t_1$. The quality of this choice is measured in terms of a performance index. For example, it could be required that the $x^i(t)$ move from a prescribed initial point to a prescribed moving target $G^i(t)$ in a minimum interval of time by choosing the $u^j(t)$ from an appropriate class of controls; or it could be required that the $x^i(t)$ move to $G^i(t)$ in a finite interval of time by using $u^j(t)$ in which the energy for control is to be a minimum. Usually the performance is measured by a functional which depends on the control variables $u^j(t)$ and the controlled variables $x^i(t)$.

We shall consider the problem of existence of an optimal control. This problem has been solved for the case of linear differential equations in [2], [3], [9], and for certain nonlinear equations in [8]; however, our treatment includes these earlier results. In Section 1, Theorems 1 and 2 state conditions such that, if there exists one allowable control which does the prescribed task, an optimal control will exist. The results of Section 2 are concerned with establishing the existence of an allowable control which accomplishes the task for various forms of the differential equation system.

1. Existence of an Optimal Control

Consider the differential system

1) $$\frac{dx^i}{dt} = f^i(t, x^1, \ldots, x^n, u^1, \ldots, u^m); \qquad i = 1, 2, \ldots, n$$

where* $f^i(t, x^1, \ldots, x^n, u^1, \ldots, u^m) = f^i(t, x, u)$ together with

$$\frac{\partial f^i}{\partial x^k}(t, x, u); \qquad i, k = 1, 2, \ldots, n$$

* In vector notation $x = (x^1, \ldots, x^n)$ and $|x| = \sum_{i=1}^{n} |x^i|$.

are real continuous functions in $R^1 \times R^n \times \Omega$, where R^n is the real n-dimensional number space and Ω is a non-empty compact subset of R^m.

For each choice of the function

$$u(t) = \big(u^1(t), \ldots, u^m(t)\big) \quad \text{on} \quad -\infty < t_0 \leqq t \leqq t_1 < \infty,$$

as a measurable vector-valued function with a graph in $\Omega \subset R^m$, the differential system

$$1)_{u(t)} \qquad \dot{x}^i = f^i\big(t, x, u(t)\big) \qquad i = 1, 2, \ldots, n$$

has a unique absolutely continuous solution, or response, $x(t)$ on a subinterval of $t_0 \leqq t \leqq t_1$, with a prescribed initial point $x_0 = x(t_0)$. This is the conclusion of the CARATHÉODORY existence theorem for differential systems [4]. Note that $u(t)$ is measurable if and only if every component $u^1(t), \ldots, u^m(t)$ is a real-valued (Lebesgue) measurable function. The response $x(t)$ is continuous, and it has a derivative, except on a set of measure zero, such that the differential system $1)_{u(t)}$ is satisfied almost everywhere. If the finite interval $t_0 \leqq t \leqq t_1$ is degenerate so $t_0 = t_1$, then the response is just the single point $x(t_0) = x_0$.

Definition. A control (or steering function) for system 1) with prescribed non-empty compact set $\Omega \subset R^m$ and prescribed initial point $x_0 \in R^n$ is a measurable vector-valued function $u(t)$, on a finite interval $t_0 \leqq t \leqq t_1$, with $u(t) \subset \Omega$, such that the response $x(t)$ with $x(t_0) = x_0$ is also defined in R^n on $t_0 \leqq t \leqq t_1$.

We shall be interested in those controls such that the response $x(t)$ travels from the prescribed initial point $x_0 = x(t_0)$ to a given moving target $G(t)$. For each t on a given finite interval $\tau_0 \leqq t \leqq \tau_1$, we specify a non-empty compact target set $G(t) \subset R^n$. Moreover, $G(t)$ varies continuously with t. Here we use the Hausdorff metric distance between two non-empty compact subsets X and Y of R^n which is the smallest real number $d = d(X, Y)$ such that X lies in the d-neighborhood of Y and Y lies in the d-neighborhood of X, *cf.* [1]. If $G(t)$ is a point for each t, then the target is a continuous curve. If $G(t)$ is a constant compact set, we have the regulator problem where the target is fixed, *cf.* [10].

Let us give $f^0(t, x^1, \ldots, x^n, u^1, \ldots, u^m)$ as a real continuous function on $R^1 \times R^n \times \Omega$ and define the cost functional of a control $u(t)$ on $t_0 \leqq t \leqq t_1$, with response $x(t)$, by

$$C(u) = \int_{t_0}^{t_1} f^0\big(t, x(t), u(t)\big)\, dt.$$

If $f^0(t, x, u) \equiv 1$, then $C(u) = t_1 - t_0$, and the cost of a control is just the time duration over which it acts.

Definition. Given the control problem

a) $\dot{x}^i = f^i(t, x, u), \qquad i = 1, 2, \ldots, n,$

b) $\Omega \subset R^m,$

c) $x_0 \in R^n,$

d) $G(t) \subset R^n \quad \text{on} \quad \tau_0 \leqq t \leqq \tau_1,$

e) $C(u)$, the cost functional,

as above. Define $\Delta = \Delta\left(f^1(t, x, u), \ldots, f^n(t, x, u), \Omega, x_0, G(t)\right)$ as the set of all controls $u(t)$, on various subintervals $t_0 \leqq t \leqq t_1$ with $\tau_0 \leqq t_0 \leqq t_1 \leqq \tau_1$, such that

$$x(t_0) = x_0 \quad \text{and} \quad x(t_1) \in G(t_1).$$

A control $u^*(t)$ in Δ is called optimal in case

$$C(u^*) \leqq C(u)$$

for every control $u(t)$ in Δ.

We now prove the basic existence theorem for optimal controls. The examples following the theorem show that an optimal control need not exist if the hypotheses of Theorem 1 are not upheld.

Theorem 1. *Given the control problem*

a) $\dot{x}^i = f^i(t, x^1, \ldots, x^n, u^1, \ldots, u^m) = g^i(t, x) + h^i_j(t, x)\, u^j$ *for* $i = 1, \ldots, n$ *and* $j = 1, \ldots, m$ *with* $g^i(t, x)$, $h^i_j(t, x)$, *and* $\frac{\partial g^i}{\partial x^k}(t, x)$, $\frac{\partial h^i_j}{\partial x^k}(t, x)$, $k = 1, \ldots, n$, *continuous on* $R^1 \times R^n$,

b) a non-empty, convex, compact restraint set $\Omega \subset R^m$,

c) the initial point $x_0 \in R^n$,

d) the continuously moving non-empty compact target set $G(t)$ *on the finite interval* $\tau_0 \leqq t \leqq \tau_1$,

e) the cost functional

$$C(u) = \int_{t_0}^{t_1} f^0\left(t, x(t), u(t)\right) dt,$$

where $f^0(t, x, u) = g^0(t, x) + h^0_j(t, x)\, u^j$, *and* $g^0(t, x)$, *and* $h^0_j(t, x)$ *are continuous on* $R^1 \times R^n$.

Assume the set Δ *of controls with responses traveling from* x_0 *to* G, *as defined above, is such that:*

A) Δ *is non-empty,*

B) there exists a real bound $B < \infty$ *for all responses* $x(t)$ *corresponding to* Δ, *that is,* $|x(t)| < B$ *uniformly for all responses.*

Then there exists an optimal control in Δ.

Proof. Since Δ is non-empty and the corresponding responses are uniformly bounded, $\inf C(u) = m > -\infty$ for all $u \in \Delta$. Either Δ is a finite set, in which case the theorem is trivially true, or we can select a sequence of controls

$$u^{(k)}(t), \quad \text{on} \quad t_0^{(k)} \leqq t \leqq t_1^{(k)} \quad \text{from } \Delta,$$

with $C(u^{(k)})$ decreasing monotonically to m. Select a subsequence (without changing the notation) such that

$$t_0^{(k)} \to t_0^* \quad \text{and} \quad t_1^{(k)} \to t_1^*, \quad \text{monotonically.}$$

Let us take the case where

$$t_0^{(k)} \leqq t_0^* \leqq t_1^* \leqq t_1^{(k)} \quad \text{for} \quad k = 1, 2, 3, \ldots$$

and consider the other cases later. Then each $u^{(k)}(t)$ is bounded and measurable on the interval $t_0^* \leqq t \leqq t_1^*$ and thus belongs to the Hilbert space $L_2[t_0^*, t_1^*]$. We assume $t_0^* < t_1^*$; for if $t_0^* = t_1^*$, then $m = 0$ and $x_0 \in G(t_1^*)$ so every choice of control is optimal.

A closed ball in Hilbert space is weakly compact [5], and thus we can select a further subsequence $u^{(k)}(t)$ with

$$u^{(k)}(t) \to u^*(t) \quad \text{weakly in} \quad L_2[t_0^*, t_1^*].$$

We next show that

$$u^*(t) \quad \text{on} \quad t_0^* \leqq t \leqq t_1^*$$

belongs to Δ.

Now a compact convex set $\Omega \subset R^m$ is precisely the intersection of a finite or a countable number of closed half-spaces, *cf.* [1]. Let

$$a_i y^i + b \geqq 0$$

be one such closed half-space in R^m. Let M be the subset of $[t_0^*, t_1^*]$ for which $u^*(t)$ lies in $a_i y^i + b < 0$.

If M has positive measure,

$$\int_{t_0^*}^{t_1^*} [a_i u^{*i}(t) + b] \varphi_m(t)\, dt < 0,$$

where φ_m is the characteristic function of M. But

$$\lim_{k\to\infty} \int_{t_0^*}^{t_1^*} [a_i u^{(k)i}(t) + b] \varphi_m(t)\, dt = \int_{t_0^*}^{t_1^*} [a_i u^{*i}(t) + b] \varphi_m(t)\, dt.$$

This is impossible since the left members are each non-negative. Thus M has measure zero. Since there are only a countable number of closed half-spaces considered, $u^*(t)$ lies in Ω except on a set of measure zero. Redefine $u^*(t)$ on the exceptional null set of $t_0^* \leqq t \leqq t_1^*$ so that $u^*(t)$ lies everywhere in Ω.

Next consider the response for $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$. The response for

$$u^{(k)}(t) \quad \text{on} \quad t_0^{(k)} \leqq t \leqq t_1^{(k)}$$

is $x^{(k)}(t)$. Then (using vector notation)

$$x^{(k)}(t) = x_0 + \int_{t_0^{(k)}}^{t} [g(s, x^{(k)}(s)) + h(s, x^{(k)}(s))\, u^{(k)}(s)]\, ds.$$

Select a further subsequence of controls so that

$$\begin{aligned}
&\lim_{k\to\infty} x^{(k)}(t) = x^*(t) && \text{weakly on} \quad [t_0^*, t_1^*],\\
&\lim_{k\to\infty} g(t, x^{(k)}(t)) = g^*(t) && \text{weakly on} \quad [t_0^*, t_1^*],\\
&\lim_{k\to\infty} h(t, x^{(k)}(t)) = h^*(t) && \text{weakly on} \quad [t_0^*, t_1^*],\\
&\lim_{k\to\infty} h(t, x^{(k)}(t))\, u^{(k)}(t) = \varphi^*(t) && \text{weakly on} \quad [t_0^*, t_1^*].
\end{aligned}$$

Then, for each fixed t on $t_0^* \leqq t \leqq t_1^*$,

$$\lim_{k\to\infty} x^{(k)}(t) = x_0 + \lim_{k\to\infty} \int_{t_0^{(k)}}^{t_0^*} [g(s, x^{(k)}(s)) + h(s, x^{(k)}(s))\, u^{(k)}(s)]\, ds +$$

$$+ \lim_{k\to\infty} \int_{t_0^*}^{t} [g(s, x^{(k)}(s)) + h(s, x^{(k)}(s))\, u^{(k)}(s)]\, ds.$$

Thus

$$\lim_{k\to\infty} x^{(k)}(t) = x_0 + \int_{t_0^*}^{t} [g^*(s) + \varphi^*(s)]\, ds.$$

Therefore

$$\lim_{k\to\infty} x^{(k)}(t) = \hat{x}(t)$$

exists, for each fixed t. Also $\hat{x}(t)$ is absolutely continuous on $[t_0^*, t_1^*]$ and $\hat{x}(t_0^*) = x_0$.

Using the Lebesgue convergence theorem [6], we find

$$\int_{t_0^*}^{t} \hat{x}(s)\, ds = \int_{t_0^*}^{t} x^*(s)\, ds,$$

so $\hat{x}(t) = x^*(t)$ for almost all t on $[t_0^*, t_1^*]$. Changing the definition of $x^*(t)$ to be precisely $\hat{x}(t)$, we now show that $x^*(t)$ is the response to the control $u^*(t)$.

Now

$$x^*(t) = x_0 + \lim_{k\to\infty} \int_{t_0^*}^{t} \left[g\left(s, x^{(k)}(s)\right) + h\left(s, x^{(k)}(s)\right) u^{(k)}(s)\right] ds,$$

so

$$\begin{aligned} x^*(t) = x_0 &+ \int_{t_0^*}^{t} g\left(s, x^*(s)\right) ds + \\ &+ \lim_{k\to\infty} \int_{t_0^*}^{t} \left[h\left(s, x^{(k)}(s)\right) u^{(k)}(s) - h\left(s, x^*(s)\right) u^{(k)}(s) + \right. \\ &\left. + h\left(s, x^*(s)\right) u^{(k)}(s) - h\left(s, x^*(s)\right) u^*(s) + h\left(s, x^*(s)\right) u^*(s)\right] ds. \end{aligned}$$

Since $u^{(k)}(s) \in \Omega$, which is compact, and since $h\left(s, x^{(k)}(s)\right) \to h\left(s, x^*(s)\right)$ almost uniformly by EGOROFF'S theorem [6], we find

$$x^*(t) = x_0 + \int_{t_0^*}^{t} \left[g\left(s, x^*(s)\right) + h\left(s, x^*(s)\right) u^*(s)\right] ds.$$

Therefore $x^*(t)$ on $t_0^* \leqq t \leqq t_1^*$ is the response to the control $u^*(t)$.

Now

$$x^{(k)}(t_1^{(k)}) \in G(t_1^{(k)}) \quad \text{for each} \quad k = 1, 2, 3, \ldots.$$

So

$$x^*(t_1^*) = \lim_{k\to\infty} \left[x^{(k)}(t_1^*) - x^{(k)}(t_1^{(k)}) + x^{(k)}(t_1^{(k)})\right],$$

and

$$x^*(t_1^*) = \lim_{k\to\infty} \left[x^{(k)}(t_1^{(k)})\right].$$

If $x^*(t_1^*)$ were not in $G(t_1^*)$, then there would exist a neighborhood N of the compact set $G(t_1^*)$, so that $x^*(t_1^*)$ is not in the closure of N. But $G(t) < N$ for t sufficiently near t_1^*. Thus $x^{(k)}(t_1^{(k)}) \in N$ for large k and yet $x^*(t_1^*)$ is not in $\overline{N}$. This is a contradiction, and therefore $x^*(t_1^*) \in G(t_1^*)$, and the control $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$ belongs to Δ.

Now compute the cost of $u^*(t)$. Here

$$C(u^{(k)}) = \int_{t_0^{(k)}}^{t_1^{(k)}} \left[g^0\left(t, x^{(k)}(t)\right) + h_j^0\left(t, x^{(k)}(t)\right) u^{(k)j}(t)\right] dt$$

and

$$\lim_{k\to\infty} C(u^{(k)}) = \int_{t_0^*}^{t_1^*} g^0\left(t, x^*(t)\right) dt + \lim_{k\to\infty} \int_{t_0^*}^{t_1^*} h_j^0\left(t, x^{(k)}(t)\right) u^{(k)j}(t)\, dt.$$

Just as above we compute

$$\lim_{k\to\infty} C(u^{(k)}) = C(u^*) = m.$$

Therefore $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$ is an optimal control.

We return to the assumption

$$t_0^{(k)} \to t_0^*, \quad t_1^{(k)} \to t_1^* \quad \text{monotonically.}$$

Suppose we do not have

$$t_0^{(k)} \leqq t_0^* \leqq t_1^* \leqq t_1^{(k)}$$

but instead, for example,

$$t_0^{(k)} \leqq t_0^* \quad \text{and} \quad t_1^{(k)} \leqq t_1^*,$$

and the other cases can be treated similarly. Extend each control $u^{(k)}(t)$ (at least for all large k) to the interval $t_0^* \leqq t \leqq t_1^*$ by defining $u^{(k)}(t) = u_0$, a constant vector in Ω, for $t_1^{(k)} \leqq t \leqq t_1^*$. Again define the weak limit in Ω,

$$u^*(t) = \lim_{k\to\infty} u^{(k)}(t) \quad \text{on} \quad t_0^* \leqq t \leqq t_1^*.$$

We must show that each response

$$x^{(k)}(t) \quad \text{on} \quad t_0^{(k)} \leqq t \leqq t_1^{(k)}$$

can be extended to the interval $t_0^{(k)} \leqq t \leqq t_1^*$ using the extended controls. Then we shall show that $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$ is in Δ and is an optimal control.

It is easily seen that all the compact sets $G(t)$ on $\tau_0 \leqq t \leqq \tau_1$ lie within one sphere, $S(0, \varrho)$ of radius ϱ, centered at the origin. Thus, for $(t_1^* - t_1^{(k)})$ sufficiently small, each $x^{(k)}(t_1^{(k)})$ lies in $S(0, \varrho)$ and has an extended response on $t_0^{(k)} \leqq t \leqq t_1^*$ which lies in $S(0, 2\varrho)$. Also

$$\lim_{k\to\infty} \left| x^{(k)}(t_1^{(k)}) - x^{(k)}(t_1^*) \right| = 0.$$

Just as above we find that

$$\lim_{k\to\infty} x^{(k)}(t) = x^*(t)$$

at each t on $t_0^* \leqq t \leqq t_1^*$, and moreover $x^*(t)$ is the absolutely continuous response to the control $u^*(t)$, and $x^*(t_0^*) = x_0$.

Now

$$x^{(k)}(t_1^{(k)}) \in G(t_1^{(k)})$$

and

$$\lim_{k\to\infty} \left| x^*(t_1^*) - x^*(t_1^{(k)}) \right| = 0.$$

Thus

$$x^*(t_1^*) = \lim_{k\to\infty} x^*(t_1^{(k)}) = \lim_{k\to\infty} \left[x^{(k)}(t_1^{(k)}) - x^{(k)}(t_1^{(k)}) + x^*(t_1^{(k)}) \right]$$

or

$$x^*(t_1^*) = \lim_{k\to\infty} x^{(k)}(t_1^{(k)}).$$

Thus

$$x^*(t_1^*) \in G(t_1^*),$$

as required. Therefore, $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$ lies in Δ.

Finally,

$$C(u^{(k)}) = \int_{t_0^{(k)}}^{t_1^{(k)}} \left[g^0\left(t, x^{(k)}(t)\right) + h_j^0\left(t, x^{(k)}(t)\right) u^{(k)j}(t) \right] dt$$

approaches the limit m as $k \to \infty$. As above we compute

$$\lim_{k \to \infty} C(u^{(k)}) = C(u^*) = m,$$

and thus $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$ is an optimal control. Q.E.D.

Remarks. Consider $\Delta(t_0)$, the subset of Δ for which the control $u(t)$ and the response $x(t)$ initiate at a fixed t_0. If $\Delta(t_0)$ is non-empty, and if the responses $x(t)$ for control in $\Delta(t_0)$ are uniformly bounded, then there exists a control in $\Delta(t_0)$ which is optimal relative to $\Delta(t_0)$. The same applies to the set $\Delta(t_0, t_1) < \Delta(t_0)$ where the control time interval $t_0 \leqq t \leqq t_1$ is fixed. If the differential system and the cost functional integrand are time-independent, each control in Δ has the same response as some control in $\Delta(\tau_0)$, after a time translation.

If $C(u) > m > -\infty$ for all controls u in Δ, or in $\Delta(t_0)$, then one requires only the uniform bound $|x(t)| < B$ for responses $x(t)$ corresponding to controls $u(t)$ with $C(u)$ near m.

Also $f^i(t, x, u)$ need only be defined and satisfy the hypotheses of the theorem, for $\tau_0 \leqq t \leqq \tau_1$, $x \in \mathcal{O} < R^n$, and $u < \Omega < R^m$, where $\mathcal{O}$ is an open set in R^n which contains the initial point x_0, the moving target $G(t)$, and all the responses of Δ, or of $\Delta(t_0)$, in a compact subset.

The hypothesis A) of the theorem concerns the domain of controllability for the problem, as will be discussed in a later section. The hypothesis B) is satisfied if

$$|f^i(t, x, u)| < \alpha, \qquad i = 1, 2, \ldots, n$$

or if

$$\left| \frac{\partial f^i}{\partial x^k}(t, x, u) \right| < \alpha, \qquad i, k = 1, 2, \ldots, n,$$

for some real α, in $[\tau_0, \tau_1] \times R^n \times \Omega$. Thus B) is always satisfied if $g^i(t, x)$ and $h_j^i(t, x)$ are linear in x.

The following examples illustrate situations where the optimal control fails to exist or is not unique.

Example 1.

$$\dot{x} = \sin 2\pi u, \quad \dot{y} = \cos 2\pi u, \quad \dot{z} = -1 \quad \text{in } R^3.$$

The initial point is (0, 0, 1), and the target is the fixed point (0, 0, 0) on the time interval $0 \leqq t \leqq t_1 \leqq 2$. The restraint set Ω is $-1 \leqq u \leqq 1$. The cost functional is $C(u) = \int_0^{t_1} (x^2 + y^2)\, dt$, and we consider the set of controls $\Delta(0)$.

Each control $u(t)$ in $\Delta(0)$ is defined on $0 \leqq t \leqq 1$. Consider the controls $u^{(k)}(t)$ such that

$$\sin 2\pi u^{(k)}(t) = \sin 2\pi k t,$$

$$\cos 2\pi u^{(k)}(t) = \cos 2\pi k t, \quad \text{for} \quad k = 1, 2, 3, \ldots.$$

Such piecewise continuous controls are easily constructed. The corresponding responses are

$$x^{(k)}(t) = \frac{1-\cos 2\pi k t}{2\pi k}, \quad y^{(k)}(t) = \frac{\sin 2\pi k t}{2\pi k}, \quad z^{(k)}(t) = 1 - t.$$

Thus $x^{(k)}(1) = 0$, $y^{(k)}(1) = 0$, $z^{(k)}(1) = 0$. The cost for each $u^{(k)}(t)$ is computed to be

$$C(u^{(k)}) = \int_0^1 \frac{1-\cos 2\pi k t}{2\pi^2 k^2}\, dt = \frac{1}{2\pi^2 k^2}.$$

Thus

$$\lim_{k\to\infty} C(u^{(k)}) = 0,$$

and $m = 0$ is the infimum for all $C(u)$ with u in $\Delta(0)$. Yet there is no optimal $u^*(t)$ on $0 \leqq t \leqq 1$ for which the cost is

$$C(u^*) = \int_0^1 (x^{*2} + y^{*2})\, dt = 0.$$

For such an optimal control $u^*(t)$ the response needed is $x^*(t) = 0$, $y^*(t) = 0$. This implies

$$\sin 2\pi u^*(t) = 0 \quad \text{and} \quad \cos 2\pi u^*(t) = 0$$

for almost all t. But this is impossible, and hence there does not exist an optimal control for this control problem. We note that the coefficient functions of the differential equation are not linear in u, but hypotheses A) and B) of the theorem are satisfied.

Example 2.

$$\dot{x} = u_1, \quad \dot{y} = u_2, \quad \dot{z} = -1 \quad \text{in } R^3.$$

The initial point is (0, 0, 1), and the target is the fixed point (0, 0, 0) on the time interval $0 \leqq t \leqq t_1 \leqq 2$. Take Ω as the compact but non-convex circle $u_1^2 + u_2^2 = 1$ in R^2. Again take $C(u) = \int_0^{t_1} (x^2 + y^2)\, dt$, and consider the controls $\Delta(0)$. Using the controls $u_1^{(k)} = \cos 2\pi k t$, $u_2^{(k)} = \sin 2\pi k t$, $k = 1, 2, \ldots$, we find $\inf C(u) = 0$ for $u \in \Delta(0)$. Yet there is no optimal control in $\Delta(0)$ which yields a cost of zero.

Example 3.

$$\dot{x} = 1, \quad \dot{y} = -x e^y u \quad \text{in } R^2.$$

The initial point is $(-1, 0)$, and the target is the fixed point $(1, 0)$ on the time interval $0 \leqq t \leqq t_1 \leqq 2$. The restraint set Ω is $0 \leqq u \leqq 2$. The cost functional is $C(u) = \int_0^{t_1} (2 - y)\, dt = \int_{-1}^{1} (2 - y)\, dx$. Each control $u(t)$ in $\Delta(0)$ is defined on $0 \leqq t \leqq 2$ and yields a response $x(t) = t - 1$, $y(t)$.

Every response $y(t)$ satisfies the inequalities

$$0 \leqq y(x) \leqq -\ln x^2, \quad \text{for} \quad x \neq 0.$$

Each continuous response joining $(-1, 0)$ to $(1, 0)$ must lie below the curve $y = -\ln x^2$ on some interval.

Thus $C(u) > \int_{-1}^{1} (2 + \ln x^2)\, dx = 0$ for each $u(t) \in \Delta(0)$. But for $u(t) = u_\varepsilon = 2 - \varepsilon$ on $0 \leqq t \leqq 2$, we compute the response $x = t - 1$, and

$$y(x) = -\ln\left[\frac{(x^2-1)}{2}(2-\varepsilon) + 1\right].$$

The cost for such a control in $\Delta(0)$ is

$$C(u_\varepsilon) = \int_{-1}^{1} \left\{2 + \ln\left[\frac{(x^2-1)}{2}(2-\varepsilon) + 1\right]\right\} dx.$$

Thus

$$\lim_{\varepsilon \to 0} C(u_\varepsilon) = 0.$$

Hence $\inf C(u) = m = 0$ for $u(t) \in \Delta(0)$. Therefore there does not exist an optimal control in $\Delta(0)$. Here we note that hypothesis B) of the theorem does not hold.

Example 4.

$$\dot{x} = 1, \quad \dot{y} = -x u \quad \text{in } R^2.$$

The initial point is $(-1, 0)$, and the target is the fixed point $(1, 0)$ on the time interval $0 \leqq t \leqq t_1 \leqq 2$. The restraint set Ω is $-1 \leqq u \leqq 1$. The cost functional is $C(u) = \int_0^{t_1} \frac{1}{1+y^2}\, dt = \int_{-1}^{1} \frac{dx}{1+y^2}$. Each control $u(t)$ in $\Delta(0)$ is defined on $0 \leqq t \leqq 2$ and yields a response $x(t) = t - 1$, $y(t)$.

Every response $y(t)$ satisfies the inequalities

$$\frac{-(1-x^2)}{2} \leqq y(x) \leqq \frac{1-x^2}{2} \quad \text{on} \quad -1 \leqq x \leqq 1.$$

The two controls $u_+(t) = +1$ and $u_-(t) = -1$ are each optimal and achieve the minimal cost. Here an optimal control in $\Delta(0)$ exists, since the hypotheses of Theorem 1 are satisfied, but it is not unique.

We close this section with an existence theorem for Lipschitz continuous controls, which is valid even if the control u enters the coefficients $f(t, x, u)$ in a nonlinear manner.

Theorem 2. *Given the control problem*

a) $\dot{x}^i = f^i(t, x^1, \ldots, x^n, u^1, \ldots, u^m)$; $i = 1, 2, \ldots, n$ *where* $f^i(t, x, u)$ *and* $\frac{\partial f^i}{\partial x^k}(t, x, u)$; $i, k = 1, 2, \ldots, n$ *are continuous in* $R^1 \times R^n \times \Omega$,

b) a non-empty compact restraint set $\Omega \subset R^m$,

c) the initial point $x_0 \in R^n$,

d) the continuously moving non-empty compact target set $G(t) \subset R^n$ *for each* t *on the finite interval* $\tau_0 \leqq t \leqq \tau_1$,

e) the cost functional

$$C(u) = \int_{t_0}^{t_1} f^0(t, x(t), u(t))\, dt,$$

where $f^0(t, x, u)$ *is continuous in* $R^1 \times R^n \times \Omega$.

For a given positive constant A *consider the class* $\Delta(\mathrm{Lip}A) \subset \Delta$ *of controls* $u(t)$, *each continuous and satisfying a Lipschitz condition*

$$|u(t) - u(t')| \leqq A|t - t'|$$

for all pairs t, t' *on some interval* $\tau_0 \leqq t_0 \leqq t$, $t' \leqq t_1 \leqq \tau_1$. *Assume*

A) $\Delta(\mathrm{Lip}A)$ *is non-empty*

B) there exists a bound $B < \infty$ *for all responses* $x(t)$ *corresponding to controls of* $\Delta(\mathrm{Lip}A)$, *that is,* $|x(t)| < B$ *uniformly for all responses. Then there exists an optimal control* $u^*(t) \in \Delta(\mathrm{Lip}A)$, *that is,* $C(u^*) \leqq C(u)$ *for all* $u(t) \in \Delta(\mathrm{Lip}A)$.

Proof. Assume that $\Delta(\mathrm{Lip}A)$ is infinite, and define

$$\inf C(u) = m > -\infty \quad \text{for all } u(t) \in \Delta(\mathrm{Lip}A).$$

Select a sequence $u^{(k)}(t)$ on $t_0^{(k)} \leqq t \leqq t_1^{(k)}$ of control of $\Delta(\mathrm{Lip}A)$ with $C(u^{(k)})$ decreasing monotonically towards m as $k = 1, 2, 3, \ldots$ tends to infinity. Select a subsequence (still called $u^{(k)}(t)$) with

$$t_0^{(k)} \to t_0^* \quad \text{and} \quad t_1^{(k)} \to t_1^* \quad \text{monotonically.}$$

Consider first the case where

$$t_0^{(k)} \leqq t_0^* \leqq t_1^* \leqq t_1^{(k)}, \qquad k = 1, 2, 3, \ldots,$$

and again we omit the trivial subcase where $t_0^* = t_1^*$. Using ASCOLI's theorem [6], select a subsequence of these controls such that

$$\lim_{k \to \infty} u^{(k)}(t) = u^*(t)$$

uniformly on $t_0^* \leqq t \leqq t_1^*$ and $u^*(t)$ is a continuous function satisfying

$$|u^*(t) - u^*(t')| \leqq A|t - t'| \quad \text{for} \quad t_0^* \leqq t, \quad t' \leqq t_1^*.$$

We must show that $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$ lies in $\Delta(\mathrm{Lip}A)$. The graph $u^*(t) \subset \Omega$, since Ω is compact.

Consider the responses $x^{(k)}(t)$ of $u^{(k)}(t)$ on $t_0^{(k)} \leqq t \leqq t_1^{(k)}$. Here in vector notation

$$x^{(k)}(t) - x^{(p)}(t) = \int_{t_0^{(k)}}^{t} f(s, x^{(k)}(s), u^{(k)}(s))\, ds - \int_{t_0^{(p)}}^{t} f(s, x^{(p)}(s), u^{(p)}(s))\, ds$$

for $t_0^* \leqq t \leqq t_1^*$ with k and $p > k$ positive integers. Then

$$|x^{(k)}(t) - x^{(p)}(t)| \leqq \int_{t_0^{(k)}}^{t_0^*} |f(s, x^{(k)}(s), u^{(k)}(s))|\, ds + \int_{t_0^{(p)}}^{t_0^*} |f(s, x^{(p)}(s), u^{(p)}(s))|\, ds +$$
$$+ \int_{t_0^*}^{t} |f(s, x^{(k)}(s), u^{(k)}(s)) - f(s, x^{(p)}(s), u^{(p)}(s))|\, ds.$$

Take k so large that $|t_0^* - t_0^{(k)}|$ and $|t_0^* - t_0^{(p)}|$ are very small and $|u^{(k)}(t) - u^{(p)}(t)|$ is uniformly small on $t_0^* \leqq t \leqq t_1^*$. Use the uniform continuity of $f(t, x, u)$ in $\tau_0 \leqq t \leqq \tau_1$, $|x| \leqq B$, $u \in \Omega$, and compute

$$|x^{(k)}(t) - x^{(p)}(t)| \leqq \frac{\varepsilon}{2} + \int_{t_0^*}^{t} |f(s, x^{(k)}(s), u^{(k)}(s)) - f(s, x^{(k)}(s), u^{(p)}(s))|\, ds + \\ + \int_{t_0^*}^{t} |f(s, x^{(k)}(s), u^{(p)}(s)) - f(s, x^{(p)}(s), u^{(p)}(s))|\, ds$$

so

$$|x^{(k)}(t) - x^{(p)}(t)| \leqq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} + \int_{t_0^*}^{t} \left|\frac{\partial f}{\partial x}\left(s, \tilde{x}(s), u^{(p)}(s)\right)\right| \cdot |x^{(k)}(s) - x^{(p)}(s)|\, ds,$$

for an arbitrarily small $\varepsilon > 0$. Write

$$z(t) = |x^{(k)}(t) - x^{(p)}(t)| \quad \text{and} \quad \left|\frac{\partial f}{\partial x}\left(s, \tilde{x}(s), u^{(p)}(s)\right)\right| < \alpha.$$

Then

$$z(t) \leqq \varepsilon + \alpha \int_{t_0^*}^{t} z(s)\, ds.$$

This integral inequality implies

$$z(t) \leqq \varepsilon\, e^{\alpha(t_1^* - t_0^*)}.$$

Therefore CAUCHY's criterion yields

$$\lim_{k \to \infty} x^{(k)}(t) = x^*(t) \quad \text{uniformly on} \quad t_0^* \leqq t \leqq t_1^*,$$

and $x^*(t)$ is continuous on $t_0^* \leqq t \leqq t_1^*$.

Now

$$x^{(k)}(t) = x_0 + \int_{t_0^{(k)}}^{t_0^*} f\left(s, x^{(k)}(s), u^{(k)}(s)\right) ds + \int_{t_0^*}^{t} f\left(s, x^{(k)}(s), u^{(k)}(s)\right) ds$$

on $t_0^* \leqq t \leqq t_1^*$. Thus

$$x^*(t) = x_0 + \int_{t_0^*}^{t} f\left(s, x^*(s), u^*(s)\right) ds,$$

and $x^*(t)$ is the response to the control $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$. Clearly $x^*(t_0^*) = x_0$. Also

$$x^*(t_1^*) = \lim_{k \to \infty} x^{(k)}(t_1^*) = \lim_{k \to \infty} \left[x^{(k)}(t_1^{(k)}) - x^{(k)}(t_1^{(k)}) + x^{(k)}(t_1^*)\right]$$

so

$$x^*(t_1^*) = \lim_{k \to \infty} x^{(k)}(t_1^{(k)}).$$

Since

$$x^{(k)}(t_1^{(k)}) \in G(t_1^{(k)}),$$

we have

$$x^*(t_1^*) \in G(t_1^*).$$

Thus $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$ belongs to $\Delta(\mathrm{Lip}\, A)$. Furthermore

$$C(u^*) = \lim_{k \to \infty} C(u^{(k)}) = m,$$

and $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$ is optimal in $\Delta(\mathrm{Lip}\, A)$.

We return to the assumption that $t_0^{(k)} \to t_0^*$ and $t_1^{(k)} \to t_1^*$, monotonically. Suppose we have $t_0^{(k)} \leqq t_0^*$ and $t_1^{(k)} \leqq t_1^*$, for example. Extend each control $u^{(k)}(t)$ to $t_0^{(k)} \leqq t \leqq t_1^*$ by defining $u^{(k)}(t) = u^{(k)}(t_1^{(k)})$ on $t_1^{(k)} \leqq t \leqq t_1^*$. For large k we can also extend the corresponding responses $x^{(k)}(t)$ to the interval $t_0^{(k)} \leqq t \leqq t_1^*$. Again

$$\lim_{k \to \infty} u^{(k)}(t) = u^*(t) \quad \text{uniformly on} \quad t_0^* \leqq t \leqq t_1^*,$$

and $u^*(t)$ satisfies a Lipschitz condition with constant A. As above we show that

$$\lim_{k \to \infty} x^{(k)}(t) = x^*(t) \quad \text{uniformly on} \quad t_0^* \leqq t \leqq t_1^*$$

and that $x^*(t)$ is the response to $u^*(t)$. Also $x^*(t_0^*) = x_0$ and $x^*(t_1^*) \in G(t_1^*)$. Thus $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$ belongs to $\Delta(\mathrm{Lip}\, A)$. Again we compute that $u^*(t)$ on $t_0^* \leqq t \leqq t_1^*$ is optimal in $\Delta(\mathrm{Lip}\, A)$. Q.E.D.

Remark. The remarks following Theorem 1 are also applicable here.

Corollary. *Assume the hypotheses of Theorem 2 for each Lipschitz constant $p = 1, 2, 3, \ldots$, and let $u_p^*(t)$ on $t_0^{(p)} \leqq t \leqq t_1^{(p)}$ be optimal in $\Delta(\mathrm{Lip}\, p)$. In addition assume*

1) the target $G(t)$ on $\tau_0 \leqq t \leqq \tau_1$ is a fixed point G,

2) for each $s > 0$ there exists a measurable control $u_s(t)$ on $t_0^{(s)} \leqq t \leqq t_1^{(s)}$ in Δ such that $C(u_s) \leqq m + s$ where $m = \inf C(u)$ for $u \in \Delta$ and also $t_1^{(s)} \leqq \tau_1 - \varepsilon$ for some $\varepsilon > 0$, independent of s,

3) there is a uniform bound $|x(t)| < B < \infty$ for all responses $x(t)$ corresponding to measurable controls in Δ,

4) for each $s > 0$ there exists a neighborhood N of G such that each point in N can be steered to G by a C^1 control having a prescribed initial instant $t_0 < \tau_1 - \varepsilon$, a duration $\leqq \varepsilon$, and a cost $\leqq s$ (this condition is considered in Theorem 4).

Then

$$\lim_{p \to \infty} C(u_p^*) = m = \inf C(u) \quad \text{for} \quad u \in \Delta.$$

Proof. It is clear that

$$C(u_p^*) \geqq C(u_{p+1}^*) \qquad p = 1, 2, 3, \ldots,$$

so $\lim_{p \to \infty} C(u_p^*)$ exists and is not less than m. But given a measurable control $u_s(t)$ in Δ, there exists a C^∞ control which approximates $u_s(t)$ almost uniformly so that the corresponding response has an end-point in N. But then x_0 can be steered to the target point G by a control $u_A(t)$ in $\Delta(\mathrm{Lip}\, A)$ for some $A > 0$, and with $C(u_A)$ arbitrarily near to m. For $p > A$ we thus have

$$C(u_A) \geqq C(u_p^*),$$

so $\lim_{p \to \infty} C(u_p^*) = m$, as required. Q.E.D.

2. Domain of Controllability

In this section we investigate the nature of hypotheses A) of Theorem 1. For simplicity we consider the problem of steering an initial point to a fixed target point, say the origin. We still maintain that the restraint set Ω is compact, but in this section we require that the zero vector is in the interior of Ω.

Definition. The domain of controllability $\mathscr{C}$ for the differential system

$$\dot{x}^i = f^i(x^1, \ldots, x^n, u^1, \ldots, u^m) \qquad i = 1, \ldots, n$$

for a given restraint set $\Omega \subset R^m$, consists of the set of all points $x_0 \in R^n$ for which there exists a measurable control $u(t) \subset \Omega$, defined on some finite interval, steering x_0 to the origin* of R^n.

Theorem 3. *Consider*

$$\dot{x}^i = f^i(x^1, \ldots, x^n; u^1, \ldots, u^m), \qquad i = 1, \ldots, n,$$

*where*** $f^i(x, u) \in C^1$ *in* $R^n \times \Omega$. *Assume the vector* $u = 0$ *is an interior point of the compact restraint set* $\Omega \subset R^m$. *Assume*

$$f^i(0, 0) = 0, \qquad i = 1, \ldots, n.$$

If the matrix $\frac{\partial f^i}{\partial u^j}(0, 0)$ *has rank* n, *then the domain of controllability* $\mathscr{C}$ *is an open connected subset of* R^n *containing the origin.*

Proof. If the origin of R^n is an interior point of $\mathscr{C}$, then it follows from general continuity arguments that $\mathscr{C}$ is open and connected. By continuity there exists a neighborhood N of the origin in Ω such that for any measurable vector $u(t)$ on $-1 \leqq t \leqq 1$ in N the corresponding solution $x(t)$ initiating at $x_0 = 0$ will be defined in R^n for $-1 \leqq t \leqq 1$.

Let

$$\varphi^i(t, u^1, \ldots, u^m) \quad \text{on} \quad -1 \leqq t \leqq 1, \qquad i = 1, \ldots, n,$$

be the solution of the given differential system for constant $(u^1, \ldots, u^m)$ in N and with $\varphi^i(0, \ldots) = 0$. We shall prove that $\frac{\partial \varphi^i}{\partial u^j}(t_0, 0)$, for some $t_0 < 0$, has rank n. Hence the differentiable map

$$u \to x = \varphi(t_0, u)$$

carries a neighborhood of $u_0 = 0$ onto a neighborhood of $x_0 = 0$. Then there is a neighborhood of $x_0 = 0$ which can be steered to the origin in time $-t_0 > 0$, and each required control is just a constant.

Now

$$\frac{\partial \dot{\varphi}^i}{\partial t} = f^i(\varphi(t, u), u), \qquad \varphi^i(t, 0) = 0.$$

We compute

$$\frac{\partial}{\partial t}\left(\frac{\partial \varphi^i}{\partial u^j}\right) = \frac{\partial f}{\partial x^k}(\varphi, u)\frac{\partial \varphi^k}{\partial u^j} + \frac{\partial f^i}{\partial u^j}(\varphi, u).$$

For $u_0 = 0$ write the matrix

$$z(t) = \frac{\partial \varphi}{\partial u}(t, 0).$$

---

* This is sometimes referred to as null controllability.

** That is, $f(x, u) \in C^0$ in $R^n \times \Omega$ and has a C^1 extension to $R^n \times \hat{\Omega}$, where $\hat{\Omega}$ is an open set containing Ω.

Then

$$\frac{dz}{dt} = f_x(0,0)\, z + f_u(0,0), \qquad z(0) = 0.$$

Then

$$z(t) = \left[I + \frac{t}{2!} f_x + \frac{t^2}{3!} f_x^2 + \cdots\right] \qquad [t f_u].$$

For a small $|t_0|$

$$z(t_0) = \frac{\partial \varphi}{\partial u}(t_0, 0)$$

has rank n, since $f_u(0,0)$ has rank n. Q.E.D.

Corollary. *Consider a linear differential system with real constant coefficients*

$$\dot{x}^i = A_k^i\, x^k + B_j^i u^j \qquad i, k = 1, \ldots, n; \quad j = 1, \ldots, m.$$

Assume the compact restraint set $\Omega \subset R^m$ contains the zero vector in its interior. If the matrix B has rank n, the domain of controllability $\mathscr{C}$ is an open connected set in R^n containing the origin.

A more difficult problem arises when the control has fewer degrees of freedom than the response, *i.e.* when $m < n$. The next theorem considers this case.

Theorem 4. *Consider*

$$\dot{x}^i = f^i(x^1, \ldots, x^n, u^1, \ldots, u^m) \qquad i = 1, \ldots, n$$

where $f(x, u) \in C^1$ in $R^n \times \Omega$. Assume the vector $u = 0$ is an interior point of the compact restraint set $\Omega \subset R^m$. Assume

1) $f^i(0,0) = 0 \qquad i = 1, \ldots, n$

and

2) there exists a vector $v \in R^m$ such that Bv lies in no invariant subspace of A with dimension $\leqq n-1$, where $B = \frac{\partial f}{\partial u}(0,0)$ and $A = \frac{\partial f}{\partial x}(0,0)$ are real matrices.

Then the domain of controllability $\mathscr{C}$ is an open connected subset of R^n containing the origin.

Proof. Consider a neighborhood $N \subset \Omega$ of the origin in R^m and $\tau_1 > 0$ such that for each measurable vector $u(t)$ on $-\tau_1 \leqq t \leqq \tau_1$ with $u(t) \subset N$ the corresponding response $x(t)$ with $x(0) = 0$ is defined on $-\tau_1 \leqq t \leqq \tau_1$.

We shall define a family of controls in N

$$u^j(t, \xi) = w^j(t, \xi_1, \ldots, \xi_n), \qquad j = 1, \ldots, m$$

on $|t| \leqq \tau_1$ and $|\xi| < \varepsilon$ for some $\varepsilon > 0$. Take

$$u(t, \xi) = v\left[X(t, t_1)\, \xi_1 + X\left(t, \frac{t_1}{2}\right) \xi_2 + \cdots + X\left(t, \frac{t_1}{n}\right) \xi_n\right]$$

where v is the vector mentioned in hypothesis 2),

$$X(t, h) = \begin{cases} 1 & \text{on} \quad |t| \leqq h \\ 0 & \text{on} \quad |t| > h, \end{cases}$$

and t_1 is a small number with $0<|t_1|<\tau_1$. For convenience we choose this piecewise continuous function $X(t, h)$, but it would be easy to use a C^∞ function which almost uniformly approximates $X(t, h)$, for each fixed h.

Note $u(t, 0)=0$ and also

$$\frac{\partial u}{\partial \xi^k} = v X\left(t, \frac{t_1}{k}\right), \qquad k = 1, \ldots, n$$

is differentiable except at $t=t_1/k$.

Let

$$x^i(t, \xi_1, \ldots, \xi_n) = x^i(t, \xi), \qquad i = 1, \ldots, n$$

be the response corresponding to $u(t, \xi)$ with

$$x^i(0, \xi) = 0, \qquad i = 1, \ldots, n.$$

Note that $x^i(t, 0)=0$, $i=1, \ldots, n$, since $f^i(0, 0)=0$. Now $x(t, \xi)$ is continuous in $(n+1)$ variables. Also, for each fixed t, the map

$$\xi \to x(t, \xi)$$

is a differentiable map of a neighborhood of the origin of R^n into R^n with the origin fixed. If

$$z_k^i(t) = \frac{\partial x^i}{\partial \xi^k}(t, 0) \qquad i, k = 1, \ldots, n$$

is non-singular, for some fixed $t<0$, then the domain of controllability $\mathscr{C}$ is an open connected subset of R^n containing the origin, as required.

Now, in vector notation,

$$x(t, \xi) = \int_0^t f\big(x(s, \xi), u(s, \xi)\big)\, ds,$$

and so

$$\frac{\partial x}{\partial \xi}(t, 0) = \int_0^t \left[A \frac{\partial x}{\partial \xi}(s, 0) + B \frac{\partial u}{\partial \xi}(s, 0)\right] ds$$

or

$$z(t) = \int_0^t [A z(s) + B u_\xi(s, 0)]\, ds,$$

so $z(t)$ is continuous near $t=0$. Also dz/dt exists, except at $t=t_1/k$, $k=1, 2, \ldots, n$.

The k^{th} column of the matrix $z(t)$ satisfies

$$z_{(k)}(t) = \int_0^t \left[A z_{(k)}(s) + B v X\left(s, \frac{t_1}{k}\right)\right] ds$$

or $k=1, 2, \ldots, n$, where the subscript $k=1, 2, \ldots, n$ designates the column. Thus we have

$$\frac{dz_{(k)}}{dt} = A z_{(k)} + B v X\left(t, \frac{t_1}{k}\right) \qquad k = 1, \ldots, n$$

with $z_{(k)}(0)=0$. But the solution of this linear differential system is

$$z_{(k)}(t)=e^{tA}\int_0^t e^{-sA}Bv\,X\left(x,\frac{t_1}{k}\right)ds$$

or

$$z_{(k)}(t)=e^{tA}\int_0^{t_1/k} e^{-sA}Bv\,ds.$$

Thus, for $|t_1|<|t|<\tau_1$,

$$z_{(1)}(t)=e^{tA}\left[t_1 I-\frac{t_1^2}{2!}A+\frac{t_1^3}{3!}A^2-\frac{t_1^4}{4!}A^3+\cdots\right]Bv,$$

$$z_{(2)}(t)=e^{tA}\left[\frac{t_1}{2}I-\frac{(t_1/2)^2}{2!}A+\frac{(t_1/2)^3}{3!}A^2-\cdots\right]Bv,$$

$$\vdots$$

$$z_{(n)}(t)=e^{tA}\left[\frac{t_1}{n}I-\frac{(t_1/n)^2}{2!}A+\frac{(t_1/n)^3}{3!}A^2-\cdots\right]Bv.$$

We shall show the linear independence of the vectors $(e^{-tA})\,z_{(1)}(t),\ldots,(e^{-tA})\,z_{(n)}(t)$.

The Vandermonde determinant

$$\begin{vmatrix} t_1 & t_1^2 & \cdots & t_1^n \\ t_1/2 & (t_1/2)^2 & \cdots & (t_1/2)^n \\ \vdots & \vdots & & \vdots \\ t_1/n & (t_1/n)^2 & \cdots & (t_1/n)^n \end{vmatrix}\neq 0.$$

Thus it is sufficient to show the linear independence of the vectors

$$\begin{aligned} &Bv+O(t_1)\\ &ABv+O(t_1)\\ &A^2Bv+O(t_1)\\ &\vdots\\ &A^{n-1}Bv+O(t_1), \end{aligned}$$

where $\lim_{t_1\to 0}O(t_1)=0$. Since the vectors $Bv, ABv,\ldots,A^{n-1}Bv$ are linearly independent, we conclude that the vectors $z_{(1)}(t), z_{(2)}(t),\ldots,z_{(n)}(t)$ are independent and that the matrix $z(t)$ is non-singular. Q.E.D.

Corollary. *Consider*

$$\dot{x}^i=f^i(x^1,\ldots,x^n,u^1,\ldots,u^m),\qquad i=1,\ldots,n$$

where $f(x,u)\in C^1$ in $R^n\times\Omega$ and the origin of R^m is an interior point of the compact set Ω. Assume

1) $f^i(0,0)=0,\qquad i=1,\ldots,n,$

2) there exists a vector $v\in R^m$ such that Bv lies in no invariant subspace of A with dimension $\leqq n-1$.

Assume also that

$$\dot{x}^i=f^i(x^1,\ldots,x^n,0,\ldots,0)\qquad i=1,\ldots,n$$

is globally asymptotically stable towards the origin of R^n. Then the domain of controllability $\mathfrak{C}$ is the entire space R^n.*

* See reference [12].

Corollary. *Consider*

$$x^{(n)} + a_{n-1}(x, x', \ldots, x^{(n-1)})\, x^{(n-1)} + \cdots + a_0(x, x', \ldots, x^{(n-1)})\, x = u$$

where the coefficients $a_i(x_1, x_2, \ldots, x_n) \in C^1$ *in* R^n *and* $|u| \leqq \varepsilon$, *for some* $\varepsilon > 0$. *We write the corresponding first order system*

$$S) \qquad \begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = x_3, \\ \quad \vdots \\ \dot{x}_n = -a_{n-1}(x_1, \ldots, x_n)\, x_n - \cdots - a_0(x_1, \ldots, x_n)\, x_1 + u. \end{cases}$$

If the system

$$\begin{aligned} \dot{x}_1 &= x_2, \\ &\vdots \\ \dot{x}_n &= -a_{n-1}(x_1, \ldots, x_n)\, x_n - \cdots - a_0(x_1, \ldots, x_n)\, x_1 \end{aligned}$$

is globally asymptotically stable towards the origin, then for S) hypotheses 1) and 2) of the theorem automatically hold and the domain of controllability for S) is all R^n.

Corollary. *Consider*

$$\dot{x}^i = f^i(x^1, \ldots, x^n, u^1, \ldots, u^m), \qquad i = 1, 2, \ldots, n$$

where $f(x, u) \in C^1$ *for* $x \in R^n$ *and* $u \in U \subset R^m$. *Assume that for each point* $x_0 \in R^n$ *there exists an interior point* $u(x_0)$ *of* U *such that*

1) $f^i(x_0, u(x_0)) = 0, \qquad i = 1, \ldots, n$

and

2) there exists a vector $v \in R^m$ *such that* Bv *lies in no invariant subspace of* A *with dimension* $\leqq n-1$, *where* $B = \frac{\partial f}{\partial u}(x_0, u(x_0))$ *and* $A = \frac{\partial f}{\partial x}(x_0, u(x_0))$ *are real matrices. Then for each pair of points* x_1 *and* x_2 *of* R^n *there exists a piecewise continuous controller* $u(t)$ *on* $t_1 \leqq t \leqq t_2$ *in* U *which steers the response* $x(t)$ *from* $x(t_1) = x_1$ *to* $x(t_2) = x_2$.

Proof. In the proof of Theorem 4 we note that there is a neighborhood of the origin which consists of points that can be steered to the origin by a piecewise continuous controller in a finite time. Also there is a neighborhood of the origin consisting of points to which the origin can be steered by a piecewise continuous controller in a finite time.

The same properties hold for each point $\bar{x}$ in R^n, after translating both x and u as required.

Now let S be the subset of R^n consisting of points to which x_1 can be steered by a piecewise continuous controller in a finite time interval. Clearly S is both open and closed in R^n. Thus $S = R^n$ and x_1 can be steered to x_2, as required, *cf.* [7]. Q.E.D.

Remarks. In the previous corollaries we can replace the hypothesis that $\dot{x}^i = f^i(x^1, \ldots, x^n, u^1, \ldots, u^m)$ be asymptotically stable for $u^j \equiv 0$, $j = 1, \ldots, m$, by the hypothesis that $\dot{x}^i = f^i(x^1, \ldots, x^n, u^1, \ldots, u^m)$ be asymptotically stable for

some appropriate choice of $u^j = u^j(x^1, \ldots, x^n)$. As a particular choice of $u^j = u^j(x)$ to satisfy the condition of asymptotic stability consider a Lyapunov function $V(x) \geqq 0$, $V(x) \in C^1$, with the choice of $u \in \Omega$ which minimizes $\left[\frac{\partial V}{\partial x^i} f^i(x, u)\right]$. If this minimum is negative definite in $0 < |x| < \infty$ for some $V(x) > 0$, with $V(x) \to \infty$ as $|x| \to \infty$, then the domain of controllability $\mathscr{C}$ is all of R^n.

We next consider the domain of controllability for certain second order differential equations

$$\ddot{x} + f(x, \dot{x}, u) = 0$$

which we write as the first order system

$$\dot{x} = y, \quad \dot{y} = -f(x, y, u)$$

in the phase plane.

Theorem 5. *Consider the differential equation*

$$\ddot{x} + f(x, \dot{x}, u) = 0$$

with $f(x, y, u) \in C^1$ for all (x, y) and $u \in \Omega$, where Ω is a compact interval containing zero as an interior point. Assume

a) $f(0, 0, 0) = 0$

b) $f_x(x, y, 0) > 0$ *and* $f_y(x, y, 0) > 0$ *everywhere*

c) $f_u(0, 0, 0) \neq 0$.

Then the domain of controllability $\mathscr{C}$ is the entire (x, y)-plane.

Proof. The system

$$\dot{x} = y$$

$$\dot{y} = -f(x, y, 0)$$

has the variable Jacobian matrix $J = \begin{pmatrix} 0 & 1 \\ -f_x & -f_y \end{pmatrix}$. Since the eigenvalues of J have negative real parts for all (x, y), the system is globally asymptotically stable to the origin, *cf.* [*12*]. However the solutions will not reach the origin in a finite time.

Now

$$A = \begin{pmatrix} 0 & 1 \\ -f_x(0, 0, 0) & -f_y(0, 0, 0) \end{pmatrix}$$

and

$$B = \begin{pmatrix} 0 \\ -f_u(0, 0, 0) \end{pmatrix}.$$

Take $v = 1$ as a 1-vector; then

$$Bv = \begin{pmatrix} 0 \\ -f_u(0, 0, 0) \end{pmatrix} \quad \text{and} \quad ABv = \begin{pmatrix} -f_u(0, 0, 0) \\ f_u \cdot f_y \end{pmatrix}$$

are linearly independent. Thus the domain of controllability to the origin is an open set. Therefore $\mathscr{C}$ is the entire (x, y)-plane. Q.E.D.

Corollary. *Consider* $\ddot{x}+f(\dot{x})+g(x)=u$

where $f(y)$ *and* $g(x)\in C^1$ *for all* (x, y) *and* $|u|\leqq\varepsilon$, *for some* $\varepsilon>0$. *Assume*

a) $f(0)=g(0)=0$

b) $f'(y)>0$ *and* $g'(x)>0$ *everywhere.*

Then the domain of controllability $\mathscr{C}$ *is the entire* (x, y)*-plane.*

We now turn from the asymptotically stable case, where the domain of controllability $\mathscr{C}$ is the entire space, to the conditionally unstable saddle-point case, where $\mathscr{C}$ is a strip in the plane.

Lemma 1. *Consider*

$\mathscr{D})$ $$\ddot{x}+f(\dot{x})-g(x)=0$$

with $f(y)$, $g(x)\in C^1$ *and* $f(0)=g(0)=0$, $f'(y)>0$, $g'(x)>0$ *for all* (x, y) *in* R^2. *Then the solution curve family of* $\mathscr{D})$ *in the* $(x, y=\dot{x})$ *phase plane is topologically equivalent to the solution curve family of the linear equation*

$$\ddot{x}+\dot{x}-x=0.$$

Proof. Write the equation $\mathscr{D})$ in the phase plane as

$\mathscr{S})$ $$\begin{cases}\dot{x}=y,\\ \dot{y}=g(x)-f(y).\end{cases}$$

We seek a homeomorphism of R^2 onto R^2 which carries the solution curves (sensed but not parametrized) of $\mathscr{S})$ onto those of

$\mathscr{L})$ $$\begin{cases}\dot{x}=y,\\ \dot{y}=x-y.\end{cases}$$

In $\mathscr{S})$ the origin is the unique critical point. The variational equation o $\mathscr{S})$ at the origin is

$$\begin{cases}\dot{x}=y,\\ \dot{y}=g'(0)\,x-f'(0)\,y,\end{cases}$$

with real eigenvalues of opposite signs. Thus $\mathscr{S})$ and $\mathscr{L})$ are topologically equivalent near the origin.

Thus $\mathscr{S})$ has exactly four solution curves which have the origin as a limit point and these approach the origin with the same directions as for $\mathscr{L})$, that is, one of these solutions of $\mathscr{S})$ lies, near the origin, in each of the four quadrants. Call these solutions of $\mathscr{S})$, which will be shown to be separatrices in the sense of MARKUS [*13*], by the numerals I, II, III, IV corresponding to the quadrants in which they approach the origin. By examining $\mathscr{S})$ on each of the coordinate axes it is easy to see that each of I, II, III, IV lies entirely in the corresponding quadrant.

Along I we have $$\frac{dy}{dx}=\frac{g(x)-f(y)}{y}$$

which is bounded from above on each compact x-interval. Hence I is single-valued over the entire positive x-axis. A similar argument shows that III is

single-valued over the entire negative x-axis. The separatrices II and IV are each single-valued over a segment of the x-axis, and they extend to infinity, that is, $x^2+y^2\to\infty$ on II and IV.

Consider a solution curve of $\mathscr{S})$ in the plane sector bounded by separatrices II and I. Here $\dot{x}=y>0$ and, in the second quadrant $\dot{y}<0$. Thus each solution curve in this sector must intersect the first quadrant and must exist in the first quadrant for all larger values of x to the right of any such intersection point. Hence the solution curves in the sector between II and I are linearly ordered (using the inclusion relation for the regions above the curves), and they form a parallel family, as described by MARKUS [*13*]. A similar analysis holds for the sector between III and IV.

In the sector bounded by IV and I each solution of $\mathscr{S})$ intersects the positive x-axis, which thereby serves as a transversal for the solutions in this sector. A similar analysis holds in the sector bounded by II and III.

Thus the only separatrices of $\mathscr{S})$ are the critical point at the origin and the solution curves I, II, III, and IV. Therefore by the general theory of separatrix configurations [*13*] we find that $\mathscr{S})$ is homeomorphic with $\mathscr{L})$, as required. Q.E.D.

Remark. Call the solution curves II and IV, leading towards the critical point of $\mathscr{S})$ with negative slope, the principal separatrices, and call I and III the minor separatrices of $\mathscr{S})$. The principal separatrices of $\mathscr{S})$, together with the critical point at the origin, form a topological image of a line in R^2 which is single-valued over the entire y-axis. The same holds for the minor separatrices of $\mathscr{S})$ over the entire x-axis.

Lemma 2. *Consider the two differential systems*

$$\mathscr{S}_\pm) \qquad \begin{cases} \dot{x}=y, \\ \dot{y}=g_\pm(x)-f(y) \end{cases}$$

where $f(y)$ *and* $g_\pm(x)=g(x)\pm C_\pm$, *for positive constants* $C_\pm$, *are in* C^1 *in* R^2. *Assume* $f(0)=g(0)=0$, $f'(y)>0$, $g'(x)>0$ *everywhere in* R^2, *and* $|g(x)|\to\infty$ *as* $|x|\to\infty$.

Then each of $\mathscr{S}_+)$ *and* $\mathscr{S}_-)$ *is topologically equivalent to the linear system, in the phase plane* $(x,\ y=\dot{x})$

$$\mathscr{L}) \qquad \begin{cases} \dot{x}=y, \\ \dot{y}=x-y. \end{cases}$$

The principal separatrices $\mathrm{II}_\pm$ *and* $\mathrm{IV}_\pm$, *together with the critical point, of* $\mathscr{S}_\pm)$ *form the topological image of a line which separates the plane in two. At each fixed ordinate the principal separatrix or critical point of* $\mathscr{S}_-)$ *lies to the right of the corresponding principal separatrix or critical point of* $\mathscr{S}_+)$. *The open band* B *between the principal separatrices and critical points of* $\mathscr{S}_+)$ *and* $\mathscr{S}_-)$ *is homeomorphic to an open band between parallel straight lines in the plane. Also each minor separatrix of* $\mathscr{S}_\pm)$ *intersects the principal separatrices of* $\mathscr{S}_\mp)$ *in at most one point, respectively* (*see Figure* 1).

Proof. By Lemma 1 the topological configuration of the solutions of $\mathscr{S}_+)$ or of $\mathscr{S}_-)$ is that of $\mathscr{L})$. It is clear that the critical point of $\mathscr{S}_-)$ lies to the right of the critical point of $\mathscr{S}_+)$, on the x-axis. By the preceeding remark the principal

separatrices II_- and IV_-, together with the critical point of $\mathscr{S}_-$), form the topological image of a line which is a single-valued covering of the y-axis. The same holds for $\mathscr{S}_+$).

Suppose II_+ and II_- intersect. At the intersection point which is furthest to the right the slope of II_- is less than that of II_+, and this contradicts the supposition of an intersection. Similarly IV_+ and IV_- cannot intersect.

Therefore the open band B bounded by the principal separatrices and critical points of $\mathscr{S}_+$) and $\mathscr{S}_-$) is homeomorphic to the open band between two parallel straight lines in the plane. In fact, the homeomorphism can be extended to the closed bands.

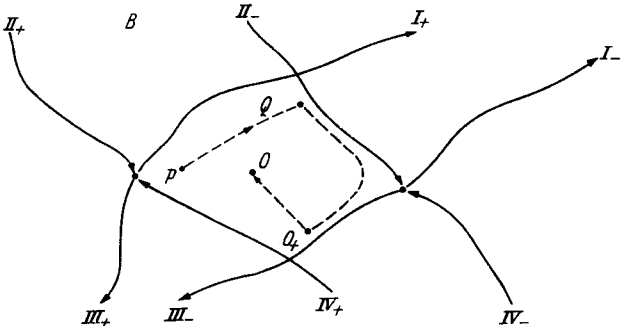

Fig. 1

Now on a minor separatrix, say I_+, we have $\dot{x}>0$ and $\dot{y}>0$. Hence I_+ intersects II_- in exactly one point. Similarly III_- intersects IV_+ in exactly one point. Q.E.D.

Theorem 6. *Consider*

$$\mathscr{D}) \qquad \ddot{x}+f(\dot{x})-g(x)=u,$$

where $f(y), g(x)\in C^1$ *for all* (x, y) *in* R^2 *and* $-C_-\leqq u\leqq C_+$ *is the finite real interval* Ω *with* $-C_-<0<C_+$. *Assume* $f(0)=g(0)=0$, $f'(y)>0$, $g'(x)>0$ *in* R^2 *and* $|g(x)|\to\infty$ *as* $|x|\to\infty$. *Then the domain of controllability* $\mathscr{C}$ *in the phase plane* $(x, y=\dot{x})$ *is precisely the open topological band* B *bounded by the principal separatrices, and critical points, of the two systems*

$$\mathscr{S}_\pm) \qquad \begin{cases} \dot{x}=y, \\ \dot{y}=g(x)\pm(C_\pm)-f(y). \end{cases}$$

Proof. Consider the open quadrilateral Q in R^2 bounded by I_+, IV_+ and II_-, III_-. The origin 0 lies in Q, and the solution $\mathscr{S}_+$) through 0 must pass through a point 0_+ in Q very near III_- before it comes to 0.

Now take an initial point p in Q. Steer p by the solution of $\mathscr{S}_+$) until it almost reaches the intersection of this solution with II_-. Then follow the solution of $\mathscr{S}_-$) around the inside of the boundary of Q to the point 0_+. Then switch to the solution of $\mathscr{S}_+$) which steers the point to 0.

If p lies in the band B but above the quadrilateral Q, start out along a solution of $\mathscr{S}_+$) until one almost reaches the intersection with II_-. Then switch to a solution of $\mathscr{S}_-$) and follow this until the point 0_+ is again reached. Then follow the solution of $\mathscr{S}_+$) into 0.

If p lies in the band B but below the quadrilateral Q, start out along a solution of $\mathscr{S}_-$) until one almost reaches IV_+. Then follow $\mathscr{S}_+$) into Q, and proceed as above.

Now let Z be a point of $R^2 - B$. If Z lies in $y>0$ and to the right of the critical point of $\mathscr{S}_-$), then Z cannot be steered out of this quadrant by a controller $-C_- \leqq u(t) \leqq C_+$ and thus Z cannot be steered to 0. Now Z follows a solution curve of

$$\begin{cases} \dot{x} = y, \\ \dot{y} = g_-(x) - f(y) + \varepsilon(t) \end{cases}$$

where $\varepsilon(t) = u(t) + C_- \geqq 0$ is measurable on some finite time interval. Thus if Z lies in the open sector bounded by II_- and I_-, then Z must lie above a solution curve of $\mathscr{S}_-$), and hence, using the parallel structure of the solution curve family in this sector, we see that Z cannot be steered to 0. If Z lies in the open sector bounded by IV_- and I_-, then Z must enter the half-plane $y>0$ to the right of the critical point of $\mathscr{S}_-$), whatever the controller $u(t)$. Therefore, if Z lies to the right of the closed band $\bar{B}$, then Z cannot be steered to the origin 0 by a measurable controller $u(t)$ with $-C_- \leqq u(t) \leqq C_+$. A similar argument shows that if Z lies to the right of the closed band B, it cannot be steered to the origin 0.

By Theorem 4 the domain of controllability $\mathscr{C}$ is an open plane set. Therefore $\mathscr{C} = B$. Q.E.D.

Remark. Each initial point in $\mathscr{C} = B$ can be steered to the origin 0 using only the solutions of $\mathscr{S}_+$) and $\mathscr{S}_-$). It is necessary to use only two switches between these two systems for each point in $\mathscr{C}$ (even though the discussion in the theorem uses three switches in certain cases).

As a final example consider the van der Pol equation

$$\ddot{x} + \mu(x^2 - 1)\,\dot{x} + x = 0$$

for a positive constant μ. In the phase plane this is the system

$$\begin{cases} \dot{x} = y \\ \dot{y} = \mu(1 - x^2)\,y - x. \end{cases}$$

The origin 0 is the unique critical point and is an unstable focus, or node. There is a unique periodic solutions S_μ which is orbitally asymptotically stable and which lies between the abscissas $x = \pm d_\mu$ (for estimates of d_μ see [*14*]).

Theorem 7. *Consider the differential system*

$$\mathscr{S}_\mu) \qquad \begin{cases} \dot{x} = y, \\ \dot{y} = \mu[1 - (x-u)^2]\,y - (x-u) \end{cases}$$

for a fixed $\mu>0$ and measurable controls $u(t)$ defined on finite intervals with $|u(t)| \leqq \varepsilon$, for some $\varepsilon>0$. Then the domain of controllability $\mathscr{C}$ to the origin 0 is all R^2 whenever

$$\varepsilon > d_\mu/2.$$

Proof. For each choice of a constant u on $|u| \leqq \varepsilon$ the corresponding system $\mathscr{S}_\mu$) has a unique critical point 0_μ, which is unstable, and a unique periodic solution, which is orbitally asymptotically stable and which lies between the abscissas $x = u \pm d_\mu$.

If the constant u satisfies $|u| < d_\mu$, then it is easy to see that 0_μ belongs to $\mathscr{C}$. If $|u| < d_\mu$ and if $\mathscr{C}$ intersects the periodic solution S_μ, then $\mathscr{C} = R^2$.

Now if $\varepsilon > d_\mu/2$, we can choose a constant u so $|u| < \varepsilon$ and yet $|2u| > d_\mu$. Then $\mathscr{C}$ intersects S_μ, and hence $\mathscr{C} = R^2$. Q.E.D.

The work of L. MARKUS was supported by NSF grant 11287 and OOR contract DA-11-022-ORD-3369.

References

[1] ALEXANDROFF, P., & H. HOPF: Topology. Berlin: Springer 1935.

[2] BELLMAN, R., I. GLICKSBERG & O. GROSS: On the bang-bang control problem. Quart. J. Appl. Math. **14**, 11—18 (1956).

[3] BOLTANSKII, V. G., R. V. GAMKRELIDZE & L. S. PONTRIAGIN: The theory of optimal processes (I — The maximum principle). Izvestiya Akademii Nauk. SSSR. Seriya Mathematicheskaya **24**, 3—42 (1960).

[4] CODDINGTON, E. A., & N. L. LEVINSON: Theory of Ordinary Differential Equations. New York: McGraw-Hill Book Co. 1955.

[5] DUNFORD, N., & J. SCHWARTZ: Linear Operators. New York: Interscience Publishers 1959.

[6] GRAVES, L. M.: The Theory of Functions of Real Variables. New York: McGraw-Hill Book Company 1946.

[7] KALMAN, R. E.: On the general theory of control systems. International Federation of Automatic Control Congress **4**, 2020—2030 (1960).

[8] KRASOVSKII, N. N.: On a problem of optimum control of nonlinear systems. P.M.M. **23**, No. 2, 209—229 (1959).

[9] LA SALLE, J. P.: Time optimal control systems. Proc. Nat. Acad. Sci. **45**, 573—577 (1959).

[10] LEE, E. B.: Methods of Optimum Feedback Control. University of Minnesota thesis, Minneapolis, Minnesota 1960.

[11] MARKUS, L., & E. B. LEE: On the existence of optimal controls. Trans. ASME J. of Basic Engineering (to appear).

[12] MARKUS, L., & H. YAMABE: Global stability criteria for differential systems. Osaka J. of Math. 1961.

[13] MARKUS, L.: Global structure of ordinary differential equations in the plane. Trans. Amer. Math. Soc. **76**, No. 1, 127—148 (1954).

[14] URABE, M., H. YANAGIARA & Y. SHINOHARA: Periodic solutions of van der Pol's equation with damping coefficient 2∼10. J. of Science of Hiroshima University 1960.

Mathematics Department
Institute of Technology
University of Minnesota
Minneapolis

(Received March 6, 1961)