

144

NGL 34-002-032
(See p. III.)

NORTH CAROLINA STATE UNIVERSITY
AT RALEIGH, DEPARTMENT OF MATHEMATICS
APPLIED MATHEMATICS RESEARCH GROUP

DR

(NASA-CR-112040) A TRANSFORMATION APPROACH
TO OPTIMAL CONTROL PROBLEMS WITH BOUNDED
STATE VARIABLES Ph.D. Thesis L.M. Hanafy
(North Carolina State Univ.) 1971 66 p
CSCL 12A

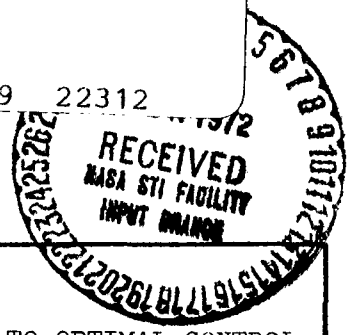
N72-20525

G3/19 22312

CR-112040
(NASA CR OR TMX OR AD NUMBER)

(CATEGORY)

FACILITY FORM 402



A TRANSFORMATION APPROACH TO OPTIMAL CONTROL
PROBLEMS WITH BOUNDED STATE VARIABLES

by

LAWRENCE MICHAEL HANAFY

Reproduced by
**NATIONAL TECHNICAL
INFORMATION SERVICE**
U S Department of Commerce
Springfield VA 22151

A TRANSFORMATION APPROACH TO OPTIMAL CONTROL
PROBLEMS WITH BOUNDED STATE VARIABLES

by

LAWRENCE MICHAEL HANAFY

A thesis submitted to the Graduate Faculty of
North Carolina State University at Raleigh
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

DEPARTMENT OF MATHEMATICS

RALEIGH

1971

APPROVED BY:

_____	<i>I</i>	_____
_____		_____
		Chairman

BIOGRAPHY

The author was born [REDACTED]

He graduated from De La Salle High School there where he was salutatorian of his senior class.

He received the Bachelor of Science degree with a major in mathematics and a minor in physics from Tulane University in 1963 where he was elected to Phi Beta Kappa. He then attended the University of California at Berkeley on a Woodrow Wilson fellowship where he received the Master of Arts degree in mathematics in 1965. From July, 1965 until August, 1968, he was employed as a research engineer with the Boeing Company in Huntsville, Alabama.

In September, 1968, he received an appointment as a research assistant in the mathematics department of North Carolina State University where for the past three years he has conducted full time study towards a doctorate.

The author married Miss Billie Jean [REDACTED] in 1964 and they have one child, Michael Thomas, two years old.

ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to all persons who have contributed inspiration, counsel and assistance in the preparation of this study. To Hans Sagan, chairman of his advisory committee, the author is particularly grateful. His guidance and advice have been invaluable, and his encouragement and friendship have made the development of this thesis a distinct pleasure. Appreciation is also extended to the other members of the committee including Doctors R. A. Struble, W. G. Dotson, Jr. and C. J. Maday for their constructive criticism and suggestions.

In addition, the author offers his thanks to the typist, Miss Marla Christopher, for perservering in fine fashion with such a tedious task.

A grant¹ from the National Aeronautics and Space Administration to North Carolina State University has provided financial support throughout this study.

Finally, the last three years of graduate study and research have been a rewarding and enjoyable experience, largely due to the patience and self sacrifice of his wife, Billie. Therefore, the author wishes to thank her and his son, Michael, whose presence makes the completion of this study a meaningful accomplishment.

¹NASA Grant NGL 34-002-032

TABLE OF CONTENTS

	Page
1. INTRODUCTION	1
2. BOUNDED STATE PROBLEMS AND THE QUESTION OF EQUIVALENCE . . .	5
2.1 The Problem	5
2.2 Review of Literature	6
2.3 Transformation into a Lagrange Problem	8
2.4 The Control and State Regions as Right Parallelepipeds	12
3. NECESSARY CONDITIONS FOR THE NONLINEAR BOUNDED STATE PROBLEM	16
4. EXAMPLE OF A LINEAR PROBLEM WITH BOUNDED STATE VARIABLES . .	33
5. THE LINEAR TIME OPTIMAL CONTROL PROBLEM WITH BOUNDED STATE VARIABLES	40
5.1 Subarcs Interior to the State Region	40
5.2 Subarcs Along the Boundary of the State Region	47
6. SUMMARY AND CONCLUSIONS	56
7. LIST OF REFERENCES	59

1. INTRODUCTION

There are several ways of approaching an optimal control problem by means of the Calculus of Variations. Perhaps the most well known technique, described in a paper by Berkovitz (1), involves adjoining additional variables to the system, commonly called slack variables, in order to transform inequality constraints into differential equation constraints. In the new problem all variables are unrestricted; hence, the classical theory can be applied. Necessary conditions for the Bolza problem are then translated into necessary conditions for optimal control. These conditions include the maximum principle of Pontryagin (20) which is seen to be a consequence of the Euler Lagrange equations and the Weierstrass necessary condition. In a paper (12) and in his book (13), Hestenes used a similar method to study optimal control problems. This method of slack variables has been widely used by a number of authors for some time. In fact, as early as 1937, F. A. Valentine (26) applied it to Lagrange problems with differential inequalities as added side conditions.

Later Berkovitz (2) applied this technique to problems with bounded state variables. He obtains essentially the results of Gamkrelidze in Chapter VI of (20) for the nonlinear problem with inequality constraints involving only state variables. Hestenes' results were extended by Russak (21) and Guinn (10) to include the bounded state case.

Another approach, described by Kalman (15), uses the Hamiltonian theory of the Calculus of Variations as it was developed by

Caratheodory (5). Sagan also utilizes this method in his book (23) where the maximum principle is shown to follow from Caratheodory's lemma in a rather simple way. However, as is pointed out by the author, the usefulness of this technique is quite limited due to the fact that unnecessary assumptions must be made regarding the differentiability of Hamilton's characteristic function and the existence of an admissible set of inception.

In this thesis a technique described by Park in (18) and (19) and applied to a simple problem by this author in (11) is to be utilized to study properties of solutions to various general problems in Optimal Control Theory, in particular those with bounded state variables. By means of an appropriate transformation of variables, optimal control problems are converted into Lagrange problems of the Calculus of Variations. This is accomplished by using mappings satisfying certain properties which take some euclidean space onto closed control and state regions.

Of course, the fundamental question must be considered as to under what conditions is the transformed problem equivalent to the original one. That is, one must know that solutions to the new problem lead to solutions of the old one and vice versa. This equivalence question is discussed in great generality by Park in (18); in this paper it will only be considered in relation to the particular problems investigated.

In chapter 2, the general problem and the method used are rigorously stated. The concept of equivalence is defined and a simple

equivalence criterion is stated and proven. The particular transformation to be used herein is exhibited and its use is justified. Also included is a survey of the literature concerning bounded state problems.

The non-linear control problems with the unit m -cube as control region and the unit n -cube as state region are considered in Chapter 3. Results similar to those of Berkovitz (2) are obtained as consequences of the well known necessary conditions for the problem of Lagrange. The translate to the optimal control setting of the hypothesis regarding the rank of the matrix of partials of the constraining equations with respect to the derivatives of the variables involved is obtained, and a new second order necessary condition analogous to the Clebsch condition is developed. In addition, the question of the singularity of optimal subarcs along boundaries of the state region is investigated, as is the behavior of solutions at boundary points where the optimal trajectory either enters or leaves the state boundary.

In Chapter 4 the results of Chapter 3 are applied to a simple example for which solutions are completely characterized. The problem is that of finding a solution to the differential equation $\dot{x} = u$ from some fixed starting point to the origin in minimal time under the restriction that $|u| \leq 1$ and including $|x| \leq 1$ and $|\dot{x}| \leq 1$ as state constraints. This chapter also serves as a preview to Chapter 5 in which the general linear time optimal control problem is considered both in the usual setting where the optimal trajectory is interior to the

state region and in the state constrained setting. In the unconstrained (in the state sense) case a generalization of the "bang - bang" principle is obtained, and analogous results are found for the state constrained case. This also leads to a new necessary condition, for the bounded state problem. This condition is always satisfied in the unconstrained case.

Finally, Chapter 6 summarizes the results obtained and offers some recommendations for possible extensions and generalizations.

2. BOUNDED STATE PROBLEMS AND THE QUESTION OF EQUIVALENCE

In this chapter a general optimal control problem with bounded state variables is defined and transformed into an equivalent Lagrange Problem by means of suitably chosen mappings defined on the state and control regions. Precise conditions for this equivalence will be formulated and specific mappings exhibited for particular state and control regions satisfying these conditions.

2.1 The Problem

Let Ω be a subset of R^m , euclidean m -space, and Γ a subset of R^n . The sets Ω and Γ will be referred to as the control region and state region respectively. Let x^0 and x^1 be points in Γ ; x^0 will be referred to as the initial point and x^1 the terminal point. The results herein may be easily generalized to the case where one has initial and terminal manifolds rather than single points. These, usually defined by systems of finite equations, are disjoint, closed subsets of Γ .

We now consider a differential system whose state at time t is characterized by a vector $x(t) = (x_1(t), \dots, x_n(t))$ in Γ and whose value is determined or controlled at that time by a vector $u(t) = (u_1(t), \dots, u_m(t))$ in Ω , the so-called control vector. Let the system be defined by the differential equations $\dot{x} = f(t, x, u)$ where f is assumed to be a continuously differentiable vector-valued function defined on $R \times \Gamma \times \Omega$ where R is the real line. This is a non-autonomous system of n non-linear first order differential equations in the $n + m$ unknowns x and u . Let an initial time t_0 be given and designate the

final time, which is variable, by t_1 .

In addition, let $f_0(t, x, u)$ be a continuously differentiable real valued function defined on $R \times \Gamma \times \Omega$. We will refer to f_0 as the cost function and its integral over the interval $[t_0, t_1]$ as the cost functional.

The problem to be considered is the following:

Problem I.

Find a sectionally continuous control $u(t)$ defined on $[t_0, t_1]$ for some $t_1 > t_0$ so that there exists $x(t)$, sectionally smooth, defined on $[t_0, t_1]$ such that

- (1) $\dot{x}(t) = f(t, x(t), u(t))$ for all $t \in [t_0, t_1]$ for which $\dot{x}(t)$ is defined,
- (2) $x(t) \in \Gamma$ and $u(t) \in \Omega$ for all $t \in [t_0, t_1]$,
- (3) $x(t_0) = x^0$ and $x(t_1) = x^1$,
- (4) $\int_{t_0}^{t_1} f_0(t, x(t), u(t)) dt$ is minimized.

This is an example of an optimal control problem with restricted state variables. Such a control $u(t)$ satisfying (1), (2), and (3) is called an admissible control. A control for which (1), (2), (3), and (4) is satisfied is called an optimal control, and its corresponding $x(t)$ is called an optimal trajectory.

2.2 Review of Literature

This type of problem has been considered by a number of people. The papers by Berkovitz (2), Guinn (10) and Russak (21) have already been discussed in Chapter 1. They both use the method of slack

variables to translate necessary conditions for the Bolza problem into necessary conditions for the above problem.

Other approaches have been more direct and do not utilize the Calculus of Variations. Gamkrelidze (20) adjoined the total time derivative of the state constraint to the cost functional and treated the resulting problem in the same fashion as he and Pontryagin did for the unconstrained problem. His "regularity" assumption, which also arises in this thesis but for different reasons, is made in order to insure that the control explicitly appears in the derivative of the constraint so that it directly influences his new cost functional.

In conjunction with numerical applications, Bryson et al. (4) developed techniques for avoiding this assumption in some cases by using higher order derivatives of the constraint. Results related to this are also demonstrated in a paper by Speyer and Bryson (24).

Dreyfus in (8) and (9) uses the method of dynamic programming to obtain results similar to those of Berkovitz. In fact, Berkovitz and Dreyfus compared their previous results in a joint paper (3) in 1965.

The penalty function approach is entirely different from all of the above. This method, first described by Chang in (6) and (7), involves the following. Instead of attempting a direct solution, an unconstrained problem is considered wherein the original cost functional is augmented by a non-negative penalty function which sharply increases the cost associated with trajectories which violate the state constraints. By using sequences of cost functionals involving more and more severe penalties it is to be expected in many cases that the

desired solution to the original problem will be obtained as the limit of the solution of the approximate problem. This technique has been subsequently refined and further results obtained by Russell (22) and Jacobson et al. (14).

In (16) Khrustalev discusses the very difficult question of sufficiency for constrained state problems, and finally McIntire and Paiewonsky, in an expository paper written in 1964, survey the techniques known at that time.

2.3 Transformation into a Lagrange Problem

We shall now state the method to be used herein to treat problems of the type described in 2.1 as Lagrange problems in the Calculus of Variations.

Consider the functions ψ and ϕ defined so that $\psi: R^k \rightarrow \Omega$ and $\phi: R^\ell \rightarrow \Gamma$ where k and ℓ are natural numbers. Assume that both ψ and ϕ are onto and continuously differentiable. Then if we let $x = \phi(y)$ and $u = \psi(\dot{z})$ and restate our original problem in terms of the new variables $y = (y_1, \dots, y_\ell)$ and $\dot{z} = (\dot{z}_1, \dots, \dot{z}_k)$ we obtain the following:

Problem II.

Find a sectionally continuous $\dot{z}(t)$ defined on $[t_0, t_1]$ for some $t_1 > t_0$ so that there exists $y(t)$, sectionally smooth, defined on $[t_0, t_1]$ such that

$$(1)' \quad \frac{\partial \phi}{\partial y}(y(t)) \dot{y}(t) = f(t, \phi(y(t)), \psi(\dot{z}(t))) \text{ for all } t \in [t_0, t_1]$$

for which $\dot{y}(t)$ is defined,

$$(2)' \quad y(t) \in R^\ell \text{ and } \dot{z}(t) \in R^k \text{ for all } t \in [t_0, t_1],$$

$$(3)' \quad \phi(y(t_0)) = x^0 \text{ and } \phi(y(t_1)) = x^1,$$

$$(4)' \int_{t_0}^t f_0(t, \phi(y(t)), \psi(\dot{z}(t))) dt \text{ is minimized.}$$

In (1)' $\frac{\partial \phi}{\partial y}$ is used to denote the $n \times l$ matrix of partial derivatives of the components of ϕ with respect to the components of the vector y . Notice also that this new problem is unconstrained in the sense that the new variables y and $\dot{z}$ are allowed to take on any values in R^l and R^k respectively.

The variable $\dot{z}$ which takes the place of the control u is introduced as a derivative so that the new problem will satisfy the hypotheses for the necessary conditions to the Lagrange problem which require that solutions, in this case $y(t)$ and $z(t)$, be sectionally smooth. That is, a sectionally smooth solution of problem II, $y(t)$ and $z(t)$, will lead to a sectionally continuous control $u(t) = \psi(\dot{z}(t))$ and a sectionally smooth trajectory $x(t) = \phi(y(t))$ as its counterpart in problem I.

We may now view problem II as a Lagrange problem with unknowns y and z , constraining differential equations given in (1)', boundary conditions (3)' and with (4)' giving the functional to be minimized. Hence all the well known classical theory associated with this problem may be applied in order to find solutions. A full discussion of this theory is contained in Chapter 6 of (23).

However, before we proceed farther along these lines, we must establish under what conditions is problem II actually equivalent to problem I, in the sense that a solution to problem II leads to a solution of problem I and vice versa. That is, we must be able to translate necessary and sufficient conditions for $y(t)$ and $z(t)$ in

problem II into necessary and sufficient conditions for $u(t) = \psi(\dot{z}(t))$ and $x(t) = \phi(y(t))$ as optimal controls and optimal trajectories in problem I. The next two theorems provide an answer to this question.

Definition 2.1. The mappings ψ and ϕ shall be called an acceptable pair of transformations provided the following are true:

- (1) $\psi: R^k \rightarrow \Omega$, $\phi: R^\ell \rightarrow \Gamma$ for some k and ℓ and both are onto and continuously differentiable;
- (2) for any sectionally smooth $x(t)$ defined on some interval $[t_0, t_1]$ such that $x(t) \in \Gamma$ for all t , there exists a sectionally smooth $y(t)$ defined on $[t_0, t_1]$ such that $y(t) \in E^\ell$ and $\phi(y(t)) = x(t)$ for all $t \in [t_0, t_1]$;
- (3) for any sectionally continuous $u(t)$ defined on some interval $[t_0, t_1]$ such that $u(t) \in \Omega$ for all t , there exists a sectionally continuous $\dot{z}(t)$ defined on $[t_0, t_1]$ such that $\psi(\dot{z}(t)) = u(t)$ for all $t \in [t_0, t_1]$.

Theorem 2.1. If ψ and ϕ are an acceptable pair of transformations and $\dot{z}(t)$, $y(t)$ defined on $[t_0, t_1]$ yield a solution to problem II then $u(t) = \psi(\dot{z}(t))$ and $x(t) = \phi(y(t))$ yield a solution to problem I.

Proof: We shall verify that $x(t)$ and $u(t)$ satisfy (1), (2), (3) and (4) of problem I. Since $\dot{x}(t) = \phi_y(y(t))\dot{y}(t) = f(t, \phi(y(t)), \psi(\dot{z}(t))) = f(t, x(t), u(t))$ we see that (1) is satisfied. Clearly (2) follows from (1) of definition 2.1, and $x(t_0) = \phi(y(t_0)) = x^0$, $x(t_1) = \phi(y(t_1)) = x^1$, so (3) is satisfied. Now suppose (4) were not true, then there would exist $\bar{x}(t)$, $\bar{u}(t)$ and $\bar{t}_1$ satisfying (1), (2) and (3) such that

$$\int_{t_0}^{\bar{t}_1} f_0(t, \bar{x}(t), \bar{u}(t)) dt < \int_{t_0}^{t_1} f_0(t, x(t), u(t)) dt.$$

Now applying (2) and (3) of definition 2.1 there exists $\bar{y}(t)$ and $\dot{\bar{z}}(t)$ such that $\bar{x}(t) = \phi(\bar{y}(t))$ and $\bar{u}(t) = \psi(\dot{\bar{z}}(t))$ for all $t \in [t_0, \bar{t}_1]$.

Then we have $\phi_y(\bar{y}(t)) \dot{\bar{y}}(t) = \dot{\bar{x}}(t) = f(t, \bar{x}(t), \bar{u}(t)) = f(t, \phi(\bar{y}(t)), \psi(\dot{\bar{z}}(t)))$ for all t and moreover $x^0 = \bar{x}(t_0) = \phi(\bar{y}(t_0))$, $x^1 = \bar{x}(t_1) = \phi(\bar{y}(t_1))$.

However,

$$\int_{t_0}^{\bar{t}_1} f_0(t, \phi(\bar{y}(t)), \psi(\dot{\bar{z}}(t))) dt = \int_{t_0}^{\bar{t}_1} f_0(t, \bar{x}(t), \bar{u}(t)) dt < \int_{t_0}^{t_1} f_0(t, x(t), u(t)) dt = \int_{t_0}^{t_1} f_0(t, \phi(y(t)), \psi(\dot{z}(t))) dt$$

which contradicts the fact that $y(t)$ and $\dot{z}(t)$ yield a solution for problem II. Thus (4) must be true and hence $x(t)$ and $u(t)$ yield a solution to problem I.

The next theorem is the converse of theorem 2.1.

Theorem 2.2. If ψ and ϕ are an acceptable pair of transformations and $u(t)$, $x(t)$ defined on $[t_0, t_1]$ yield a solution to problem I then any sectionally smooth $y(t)$ and $z(t)$ such that $x(t) = \phi(y(t))$ and $u(t) = \psi(\dot{z}(t))$ for $t \in [t_0, t_1]$ yield a solution to problem II.

Proof: We know that at least one such $y(t)$ and $\dot{z}(t)$ exist by (2) and (3) of definition 2.1. We must therefore show that such a $y(t)$ and $\dot{z}(t)$ satisfy (1)', (2)', (3)' and (4)' of problem II. Notice that $\phi_y(y(t)) \dot{y}(t) = \dot{x}(t) = f(t, x(t), u(t)) = f(t, \phi(y(t)), \psi(\dot{z}(t)))$; thus we have that (1)' is true. Moreover (2)' is trivially satisfied and since $\phi(y(t_0)) = x(t_0) = x^0$ and $\phi(y(t_1)) = x(t_1) = x^1$, (3)' is also. Now if $\bar{y}(t)$ and $\dot{\bar{z}}(t)$ defined on $[t_0, \bar{t}_1]$ satisfies (1)', (2)', (3)' and

$$\int_{t_0}^{\bar{t}_1} f_0(t, \phi(\bar{y}(t)), \psi(\dot{\bar{z}}(t))) dt < \int_{t_0}^{t_1} f_0(t, \phi(y(t)), \psi(\dot{z}(t))) dt$$

then it is easily seen that $\bar{x}(t) = \phi(\bar{y}(t))$ and $\bar{u}(t) = \psi(\dot{\bar{z}}(t))$ will contradict the optimality of $x(t)$ and $u(t)$. Thus $y(t)$ and $\dot{z}(t)$ must yield a solution to problem II.

The following corollary follows immediately from theorems 2.1 and 2.2.

Corollary 2.1. If ψ and ϕ are an acceptable pair of transformations then any necessary or sufficient condition for $y(t)$ and $\dot{z}(t)$ to be a solution of problem II yields a necessary or sufficient condition for $x(t)$ and $u(t)$ to be a solution of problem I when the condition is restated in terms of $x(t) = \phi(y(t))$ and $u(t) = \psi(\dot{z}(t))$.

A much more general discussion of the equivalence of minimization problems is contained in (18), particularly in regard to unconstrained problems.

2.4 The Control and State Regions as Right Parallelepipeds

Consider problem I with

$$\Omega = \{u \in R^m: a_i \leq u_i \leq b_i, i = 1, \dots, m\} \text{ and}$$

$$\Gamma = \{x \in R^n: c_i \leq x_i \leq d_i, i = 1, \dots, n\}$$

where a_i, b_i, c_i and d_i are real numbers such that $a_i < b_i$ and $c_i < d_i$ for each i . Then Ω and Γ defined in this way constitute right parallelepipeds. Now define $\psi: R^m \rightarrow \Omega$ by $\psi = (\psi_1, \dots, \psi_m)$ where

$$u_i = \psi_i(\dot{z}) = \frac{1}{2} ((b_i - a_i) \sin \dot{z}_i + (b_i + a_i))$$

for $i = 1, \dots, m$. Also let $\phi: R^n \rightarrow \Gamma$ be given by $\phi = (\phi_1, \dots, \phi_n)$ where

$$x_i = \phi_i(y) = \frac{1}{2} ((d_i - c_i) \sin y_i + (d_i + c_i))$$

for $i = 1, \dots, n$.

Theorem 2.3. For this choice of Ω , Γ , ψ and ϕ , the mappings ψ and ϕ constitute an acceptable pair of transformations.

Proof: Since $-1 \leq \sin \dot{z}_i \leq 1$ for all real numbers $\dot{z}_i$, we have

$$\frac{(a_i - b_i)}{2} \leq \frac{(b_i - a_i)}{2} \sin \dot{z}_i \leq \frac{(b_i - a_i)}{2}, \text{ and therefore}$$

$$\begin{aligned} a_i &= \frac{(a_i - b_i)}{2} + \frac{(b_i + a_i)}{2} \leq \frac{(b_i - a_i)}{2} \sin \dot{z}_i + \frac{(b_i + a_i)}{2} = \psi_i(\dot{z}) \\ &\leq \frac{(b_i - a_i)}{2} + \frac{(b_i + a_i)}{2} = b_i \end{aligned}$$

for $i = 1, \dots, n$.

Thus we see that $\psi(\dot{z}) \in \Omega$ for all $\dot{z} \in \mathbb{R}^m$. Similarly we can show that $\phi(y) \in \Gamma$ for all $y \in \mathbb{R}^n$. Moreover ψ and ϕ are clearly continuously differentiable. Now let $u(t)$ be a sectionally continuous function defined on $[t_0, t_1]$ such that $u(t) \in \Omega$ for all $t \in [t_0, t_1]$. Define $\dot{z}(t)$ by

$$\dot{z}_i(t) = \sin^{-1} \left[\frac{2u_i(t) - (b_i + a_i)}{(b_i - a_i)} \right].$$

We first notice that each $\dot{z}_i(t)$ is well defined for all $t \in [t_0, t_1]$.

This is true since $u(t) \in \Omega$ means that $a_i \leq u_i(t) \leq b_i$ for $i = 1, \dots, m$ and hence it follows that $a_i - b_i = 2a_i - (b_i + a_i) \leq 2u_i(t) - (b_i + a_i) \leq 2b_i - (b_i + a_i) = b_i - a_i$.

Dividing this inequality by $b_i - a_i$ we obtain

$$-1 \leq \frac{2u_i(t) - (b_i + a_i)}{(b_i - a_i)} \leq 1.$$

Hence the $\sin^{-1}$ of the middle term in the above inequality is well defined. Note that this argument yields that ψ is onto since $u(t)$ may in general be any point of Ω and clearly if $\dot{z}$ is defined in the above fashion for any $u \in \Omega$ we have $\psi(\dot{z}) = u$. Moreover, since $\sin^{-1}$ is a sectionally continuous function, we see that the $\dot{z}(t)$ which we have defined will be sectionally continuous also. Therefore ψ satisfies condition (3) of definition 2.1. If $x(t)$ is a sectionally smooth function defined on $[t_0, t_1]$ such that $x(t) \in \Gamma$ for all $t \in [t_0, t_1]$, we can satisfy condition (2) of definition 2.1 by defining

$$y_i(t) = \sin^{-1} \left[\frac{2x_i(t) - (d_i + c_i)}{(d_i - c_i)} \right] + 2\pi j(i, t)$$

for $i = 1, \dots, n$ and $t \in [t_0, t_1]$ where $j(i, t)$ is a integer chosen for each i and t to insure that $y_i(t)$ is not just sectionally continuous but also sectionally smooth. Since $\sin(2\pi j) = 0$ for all integers j , this term does not affect the relationship $\phi(y(t)) = x(t)$. Thus the theorem is proved and we have shown that ψ and ϕ are an acceptable pair of transformations.

We now have that for this particular choice of Ω , Γ , ψ and ϕ that theorems 2.1 and 2.2 apply, and hence, in this case, problem I is equivalent to problem II.

In the case that some of the components of u or x are to be unrestricted, while the remainder are constrained between maximum and minimum values as previously, one simply defines $\psi_i(\dot{z}) = u_i$ or $\phi_i(y) = x_i$ for those particular ones and defines the rest of the components of ψ and ϕ as is done previously with the sine function. Clearly, in this case, the results of theorem 2.3 also follow. For numerical applications the previous formulation is more desirable as in effect it covers both cases. When a component of u or x is to be unrestricted, one simply inputs to the system maximum and minimum values of the variable which are exceedingly large and exceedingly small respectively, thus effectively eliminating the constraint.

For simplicity sake, in the remainder of this text we shall let Ω be the unit m -cube and Γ be the unit n -cube. That is, we will set $a_i = -1$, $b_i = 1$ for $i = 1, \dots, m$ and $c_i = -1$, $d_i = 1$ for $i = 1, \dots, n$. Then ψ and ϕ reduce to $\psi(\dot{z}) = u = \sin \dot{z} = (\sin \dot{z}_1, \dots, \sin \dot{z}_m)$ and $\phi(y) = x = \sin y = (\sin y_1, \dots, \sin y_n)$.

Examples of acceptable choices of ψ and ϕ corresponding to more general control and state regions are contained in (18).

3. NECESSARY CONDITIONS FOR THE NONLINEAR BOUNDED STATE PROBLEM

Let us first introduce for notational convenience an operation on pairs of vectors which we shall designate by " $\star$ ". If $a = (a_i)_{i=1}^n$ and $b = (b_i)_{i=1}^n$ are n -vectors, then define $a \star b \in R^n$ to be the n -vector $(a_i b_i)_{i=1}^n$. Clearly " $\star$ " is commutative, associative and distributive with respect to addition. That is, if $a, b, c \in R^n$ then $a \star (b + c) = a \star b + a \star c$. We shall use $a \cdot b$ to designate the ordinary dot product of the vectors a and b .

In addition, for $a \in R^n$ we introduce the notation " $\text{diag}(a)$ " and define $\text{diag}(a)$ to be the $n \times n$ diagonal matrix whose diagonal elements are precisely the components of a . Notice that for $a, b \in R^n$ we have $\text{diag}(a)b = a \star b$.

Now consider problem I in section 2.1 with

$$\Omega = \{u \in R^m : |u_i| \leq 1, i = 1, \dots, m\}, \text{ the unit } m\text{-cube, and}$$

$$\Gamma = \{x \in R^n : |x_i| \leq 1, i = 1, \dots, n\}, \text{ the unit } n\text{-cube.}$$

Then if we define ϕ and ψ by

$$\phi(y) = x = \sin y \quad \text{and}$$

$$\psi(\dot{z}) = u = \sin \dot{z},$$

then theorems 2.1, 2.2, and 2.3 insure that for this case problem I is equivalent to problem II in section 2.3. Note that we now have that $\frac{\partial \phi}{\partial y} = \text{diag}(\cos y)$ so that $\frac{\partial \phi}{\partial y} \dot{y} = \cos y \star \dot{y}$. Hence the constraining differential equations for problem II become

$$\cos y \dot{y} = f(t, \sin y, \sin \dot{z}).$$

The boundary conditions are

$$\sin(y(t_0)) = x^0, \text{ and } \sin(y(t_1)) = x^1,$$

and the cost functional is

$$\int_{t_0}^{t_1} f_0(t, \sin y(t), \sin \dot{z}(t)) dt.$$

Define the functions ϕ_0, ϕ, ψ_0 and ψ_1 taking values in R^n by

$$\phi_0(t, y, \dot{z}) = f_0(t, \sin y, \sin \dot{z}),$$

$$\phi(t, y, \dot{y}, \dot{z}) = \cos y \dot{y} - f(t, \sin y, \sin \dot{z}),$$

$$\psi_0(y) = \sin y - x^0 \text{ and}$$

$$\psi_1(y) = \sin y - x^1.$$

Then, if we restate problem II of section 2.3 for the particular problem considered in this chapter using the above functions, we obtain:

Problem II'

Find piecewise smooth vector functions $y(t)$ and $z(t)$ defined on $[t_0, t_1]$ where t_0 is fixed and t_1 is variable which satisfy the constraining differential equations $\phi(t, y, \dot{y}, \dot{z}) = 0$, the initial conditions $\psi_0(y(t_0)) = 0$, the terminal conditions $\psi_1(y(t_1)) = 0$ and is such that $\int_{t_0}^{t_1} \phi_0(t, y(t), z(t)) dt$ is minimized.

The following theorem summarizes the results found in chapter 6

of Sagan's book (23) concerning properties of solutions to the above problem.

Theorem 3.1. If $y(t)$ and $z(t)$ defined on $[t_0, t_1]$ constitute a solution to problem II' and if the following conditions are satisfied

- (1) ϕ_0, ϕ, ψ_0 and ψ_1 are continuously differentiable,
- (2) along $y(t)$ and $z(t)$ we have $\text{rank}(\frac{\partial \phi}{\partial (\dot{y}, \dot{z})}) = n$ for all $t \in [t_0, t_1]$ where $\frac{\partial \phi}{\partial (\dot{y}, \dot{z})}$ is the $n \times (n + m)$ matrix of partials of the components of ϕ with respect to the components of $\dot{y}$ and $\dot{z}$;
- (3) $\text{rank}(\frac{\partial \psi_0}{\partial y}(y(t_0))) = n = \text{rank}(\frac{\partial \psi_1}{\partial y}(y(t_1)))$,

then there exists a vector function $\lambda(t) = (\lambda_1(t), \dots, \lambda_n(t))$ defined on $[t_0, t_1]$ which is continuous except possibly at the points where $y(t)$ and/or $z(t)$ are not smooth and a constant $\lambda_0 \leq 0$ with $(\lambda_0, \lambda_1(t), \dots, \lambda_n(t)) \neq (0, 0, \dots, 0)$ for all $t \in [t_0, t_1]$ such that the following statements are true:

(1) Euler Lagrange Equations

define $h(t, y, \dot{y}, \dot{z}, \lambda) = -\lambda_0 \phi_0(t, y, \dot{z}) + \lambda \cdot \phi(t, y, \dot{y}, \dot{z})$, then the differential equations

$$h_y(t, y, \dot{y}, \dot{z}, \lambda) - \frac{d}{dt} h_{\dot{y}}(t, y, \dot{y}, \dot{z}, \lambda) = 0 \quad \text{and}$$

$$\frac{d}{dt} h_{\dot{z}}(t, y, \dot{y}, \dot{z}, \lambda) = 0$$

are satisfied along every smooth arc of $y(t)$ and $z(t)$ where $h_y, h_{\dot{y}}$, and $h_{\dot{z}}$ denote the vectors of partials of the function h with respect to the vectors $y, \dot{y}$ and $\dot{z}$ respectively;

(2) Corner Conditions

across every corner of $y(t)$ and $z(t)$ we must have that each component of

$$h_{\dot{y}}(t, y(t), \dot{y}(t), \dot{z}(t), \lambda(t)),$$

$$h_{\dot{z}}(t, y(t), \dot{y}(t), \dot{z}(t), \lambda(t)), \quad \text{and}$$

$$h(t, y(t), \dot{y}(t), \dot{z}(t), \lambda(t)) - \dot{y}(t) \cdot h_{\dot{y}}(t, y(t), \dot{y}(t), \dot{z}(t), \lambda(t)) - \\ - \dot{z}(t) \cdot h_{\dot{z}}(t, y(t), \dot{y}(t), \dot{z}(t), \lambda(t))$$

are continuous;

(3) Transversality Conditions

there exists a constant vector $v = (v_1, \dots, v_n)$ such that

$$(\lambda_0, v_1, \dots, v_n) \neq (0, \dots, 0) \text{ and}$$

$$0 = -\lambda_0 \phi_0 - h_{\dot{y}} \cdot \dot{y} - h_{\dot{z}} \cdot \dot{z}$$

$$v \Psi_{0y} = h_{\dot{y}} \quad \text{and}$$

$$0 = h_{\dot{z}}$$

where the above relations are evaluated at $t_0, y(t_0), \dot{y}(t_0), \dot{z}(t_0)$

and $\lambda(t_0)$, and there exists a constant vector $\mu = (\mu_1, \dots, \mu_n)$

such that $(\lambda_0, \mu_1, \dots, \mu_n) \neq (0, \dots, 0)$ and

$$0 = -\lambda_0 \phi_0 - h_{\dot{y}} \cdot \dot{y} - h_{\dot{z}} \cdot \dot{z}$$

$$\mu \Psi_{1y} = h_{\dot{y}} \quad \text{and}$$

$$0 = h_{\dot{z}}$$

evaluated at $t_1, y(t_1), \dot{y}(t_1), \dot{z}(t_1)$ and $\lambda(t_1)$;

(4) Clebsch Condition

for each $t \in [t_0, t_1]$ and for all vectors $\sigma \in \mathbb{R}^n$, $\rho \in \mathbb{R}^m$

which are solutions to the linear systems

$$\Phi_{\dot{y}}(t, y(t), \dot{y}(t), \dot{z}(t))\sigma = 0 \quad \text{and}$$

$$\Phi_z(t, y(t), \dot{y}(t), \dot{z}(t))\rho = 0$$

we must have that

$$\sigma \cdot h_{\dot{y}\dot{y}} \sigma + \rho \cdot h_{\dot{z}\dot{y}} \sigma + \sigma \cdot h_{\dot{y}\dot{z}} \rho + \rho \cdot h_{\dot{z}\dot{z}} \rho \geq 0$$

evaluated at $t, y(t), \dot{y}(t), \dot{z}(t)$ and $\lambda(t)$ provided the functions f_0 and f are twice continuously differentiable (see Hestenes (13));

(5) Weierstrass Condition

for all $(t, y(t), \dot{y}, \dot{z})$ satisfying the constraining equations, $t \in [t_0, t_1]$, we have

$$E(t, y(t), \dot{y}(t), \dot{z}(t), \dot{y}, \dot{z}, \lambda(t)) \geq 0$$

where

$$\begin{aligned} E(t, y, \dot{y}, \dot{z}, \dot{y}, \dot{z}, \lambda) &= h(t, y, \dot{y}, \dot{z}, \lambda) \\ &- h(t, y, \dot{y}, \dot{z}, \lambda) + (\dot{y} - \dot{y}) \cdot h_{\dot{y}}(t, y, \dot{y}, \dot{z}, \lambda) \\ &+ (\dot{z} - \dot{z}) \cdot h_{\dot{z}}(t, y, \dot{y}, \dot{z}, \lambda). \end{aligned}$$

Using these necessary conditions we shall now determine what they mean in terms of our particular problem. Notice that since f_0 and f were assumed to be continuously differentiable we have that ϕ_0, ϕ, ψ_0 and ψ_1 are by definition continuously differentiable, Hence assumption (1) of theorem 3.1 is satisfied for our particular problem.

Now let $x(t)$ and $u(t)$ defined on $[t_0, t_1]$ yield a solution to problem I of section 2.1. with our particular control region Ω and state region Γ . Then by theorem 2.3 there exists $y(t)$ and $z(t)$ such

that $\sin(y(t)) = x(t)$ and $\sin(\dot{z}(t)) = u(t)$ for all $t \in [t_0, t_1]$, and moreover, theorem 2.2 tells us that $y(t)$ and $z(t)$ yield a solution to problem II'. We assume that hypotheses (2) and (3) of theorem 3.1 are satisfied. Therefore theorem 3.1 may be applied to $y(t)$ and $z(t)$ thereby establishing corresponding conditions on $u(t)$ and $x(t)$. However, before proceeding further, some results should be exhibited describing under which circumstances assumptions (2) and (3) of theorem 3.1 will be met. The next four theorems will accomplish this.

Let $M(t)$ be the $n \times m$ matrix $f_u \text{diag}(\cos \dot{z})$ evaluated at $(t, y(t), \dot{y}(t), \dot{z}(t))$. Suppose at time t there are precisely ℓ state constraints in effect; that is, ℓ components of the vector $x(t)$ have absolute value one. Define $\bar{M}(t)$ to be the $\ell \times m$ submatrix of $M(t)$ consisting only of the rows of $M(t)$ corresponding to components of $x(t)$ with absolute value one.

Theorem 3.2. For all $t \in [t_0, t_1]$, $\bar{M}(t)$ has rank ℓ .

Proof: This theorem follows from assumption (2) of theorem 3.1.

Since $\Phi = \cos y \dot{y} - f$, we see that

$$\frac{\partial \Phi}{\partial(\dot{y}, \dot{z})} = (\text{diag}(\cos y(t)), M(t)).$$

If $|x_i(t)| < 1$ then since $x_i(t) = \sin(y_i(t))$ we have that $\cos y_i(t) \neq 0$ and thus the rows of the above matrix corresponding to the $n - \ell$ components of $x(t)$ with absolute value less than one are linearly independent. However, for the ℓ coordinates where $|x_i(t)| = 1$ we have that $\cos y_i(t) = 0$ and therefore $\frac{\partial \Phi}{\partial(\dot{y}, \dot{z})}$ will have maximum rank if and only if the rows of $M(t)$ corresponding to these components are

linearly independent. In other words this means that $\bar{M}(t)$ must have rank ℓ .

We can see from the above proof that this theorem gives a necessary and sufficient condition for assumption (2) of theorem 3.1 to be satisfied. Moreover, it can be easily seen that all trajectories which never meet the state constraint boundary automatically satisfy hypothesis (2).

Theorem 3.3. If $x(t)$ is in the interior of Γ for all $t \in [t_0, t_1]$, then hypothesis (2) of theorem 3.1 is satisfied.

Proof: In this case $|x_i(t)| < 1$ for all $i = 1, \dots, n$ and $t \in [t_0, t_1]$ and so $\cos(y_i(t)) \neq 0$ for all $i = 1, \dots, n$ and $t \in [t_0, t_1]$. Therefore $\text{rank}(\text{diag}(\cos y(t))) = n$ and hence,

$$\text{rank} \frac{\partial \Phi}{\partial (\dot{y}, \dot{z})} = \text{rank}(\text{diag}(\cos y(t)), M(t)) = n.$$

The next theorem tells us something about when a system is over-constrained. It simply says that at any time, there cannot be more constraints in effect, both state and control, than there are control variables. Suppose that at time t there are k components of $u(t)$ such that $|u_i(t)| = 1$; that is, there are k control constraints in effect.

Theorem 3.4. $k + \ell \leq m$ at every time $t \in [t_0, t_1]$.

Proof: According to theorem 3.2, the $\ell \times m$ matrix $\bar{M}(t)$ must have rank ℓ . Now $M(t) = f_u \text{diag}(\cos \dot{z}(t))$ and consequently each element of column i of $M(t)$ is multiplied by $\cos \dot{z}_i(t)$ for $i = 1, \dots, m$. Moreover, each element of column i of $\bar{M}(t)$ is multiplied by $\cos \dot{z}_i(t)$ also. For each component of $u(t)$ with absolute value one we have that

$\cos \dot{z}_i(t) = 0$ since $u_i(t) = \sin(\dot{z}_i(t))$. Hence at least k columns of $\bar{M}(t)$ have all zero elements. Thus the rank of $\bar{M}(t)$ is at most $m - k$, and so

$$\text{rank } (\bar{M}(t)) = \ell \leq m - k \text{ or equivalently } k + \ell \leq m.$$

The next theorem gives a necessary and sufficient condition for hypothesis (3) of theorem 3.1 to be satisfied.

Theorem 3.5. The vectors $x(t_0)$ and $x(t_1)$ are interior to Γ .

Proof: Recall that $\Psi_0 = \sin(y) - x^0$ and $\Psi_1 = \sin(y) - x^1$. Thus we

have $\frac{\partial \Psi_0}{\partial y} = \text{diag}(\cos y) = \frac{\partial \Psi_1}{\partial y}$ and this matrix will have rank n if and only if $\cos y_i \neq 0$ for $i = 1, \dots, n$, or equivalently, $|x_i| < 1$ for $i = 1, \dots, n$. But this is the same as requiring that x be interior to Γ . Therefore $x(t)$ will satisfy hypothesis (3) of theorem 3.1 if and only if both $x(t_0)$ and $x(t_1)$ are interior to Γ .

We shall now proceed to determine the consequences of conclusions (1) through (5) of theorem 3.1 for $x(t)$ and $u(t)$.

Let the function H , called the Hamiltonian, be defined by

$$H(t, x, u, \lambda) = \lambda_0 f_0(t, x, u) + \lambda \cdot f(t, x, u).$$

Theorem 3.6. There exists a vector function $\lambda(t) = (\lambda_1(t), \dots, \lambda_n(t))$ defined on $[t_0, t_1]$ which is continuous except possibly at the points where $u(t)$ is not continuous and a constant $\lambda_0 \leq 0$ with $(\lambda_0, \lambda_1(t), \dots, \lambda_n(t)) \neq 0$ for all $t \in [t_0, t_1]$ such that for $i = 1, \dots, n$ either

$$|x_i(t)| = 1 \text{ or } \dot{\lambda}_i(t) + H_{x_i}(t, x(t), u(t), \lambda(t)) = 0$$

and for $j = 1, \dots, m$ either

$$|u_j(t)| = 1 \text{ or } H_{u_j}(t, x(t), u(t), \lambda(t)) = 0$$

on every smooth arc of $x(t)$.

Proof: First notice that

$$h = \lambda \cdot (\dot{y} \cos y) - H(t, \sin y, \sin \dot{z}, \lambda)$$

and thus $h_{\dot{y}} = \lambda \cos y$, so

$$\frac{d}{dt} h_{\dot{y}} = \dot{\lambda} \cos y - \lambda \sin y \dot{y}, \text{ and}$$

$$h_y = -\lambda \dot{y} \sin y - H_x(t, \sin y, \sin \dot{z}, \lambda) \cos y.$$

Hence conclusion (1) of theorem 3.1 becomes

$$\begin{aligned} h_y - \frac{d}{dt} h_{\dot{y}} &= -\lambda \dot{y} \sin y - H_x(t, \sin y, \sin \dot{z}, \lambda) \cos y \\ &\quad - \dot{\lambda} \cos y + \lambda \sin y \dot{y} = 0 \end{aligned}$$

That is,

$$(\dot{\lambda} + H_x(t, \sin y, \sin \dot{z}, \lambda)) \cos y = 0$$

from which the first part of the theorem follows. For the second part we have

$$h_{\dot{z}} = H_u(t, \sin y, \sin \dot{z}, \lambda) \cos \dot{z}$$

and so from the Euler Lagrange equations $H_u \cos \dot{z}$ must be constant on each smooth subarc of $x(t)$. In addition, the corner conditions

tell us that this quantity is continuous, and hence $H_u * \cos \dot{z}$ must be constant for all $t \in [t_0, t_1]$. Moreover, looking at the transversality conditions we see that $h_z = H_u * \cos \dot{z} = 0$ for $t = t_0$ and $t = t_1$. Thus, along the solution

$$H_u(t, \sin y, \sin \dot{z}, \lambda) * \cos \dot{z} \equiv 0$$

from which the second part of the theorem follows.

Corollary 3.1. If $x(t)$ is in the interior of Γ for all $t \in [t_0, t_1]$, then

$$\dot{\lambda}(t) + H_x(t, x(t), u(t), \lambda(t)) = 0$$

on every smooth subarc of $x(t)$.

Corollary 3.2. If $u(t)$ is in the interior of Ω for all $t \in [t_0, t_1]$, then

$$H_u(t, x(t), u(t), \lambda(t)) = 0$$

on every smooth subarc of $x(t)$.

Definition 3.1. Let i be an integer such that $1 \leq i \leq n$; then the point $x(t^*)$ for $t^* \in (t_0, t_1)$ is called an i-boundary corner of $x(t)$ provided either $x(t)$ is not differentiable at $t = t^*$ or t^* is a point of discontinuity of $u(t)$, and $|x_i(t^*)| = 1$.

Theorem 3.7. For each $i = 1, \dots, n$, $\lambda_i(t)$ is continuous except possibly at i -boundary corners of $x(t)$. The function $H(t, x(t), u(t), \lambda(t))$ is continuous for all $t \in [t_0, t_1]$.

Proof: From the corner conditions, since $h_y = \lambda \cos y$ we must have that the product $\lambda_i(t) \cos y_i(t)$ is continuous across each corner of $(y(t), z(t))$. Let t^* be a corner of $(y(t), z(t))$. Then either $x(t)$ is not differentiable at $t = t^*$ or t^* is a point of discontinuity of $u(t)$. Suppose t^* is not an i -boundary corner. Then $|x_i(t)| < 1$ and consequently $\cos(y_i(t^*)) \neq 0$. Thus since $\cos(y_i(t))$ is a continuous function which is non-zero for $t = t^*$ and $\lambda_i(t) \cos y_i(t)$ is continuous at $t = t^*$, we must have that $\lambda_i(t)$ is continuous at $t = t^*$. Since we already know that $\lambda_i(t)$ is continuous at all points where $(y(t), z(t))$ is smooth, the first part of the theorem follows. In addition the corner conditions yield that $h - \dot{y} \cdot h_y - \dot{z} \cdot h_z$ must be continuous across corners. But it was seen in the proof of theorem 3.6 that along the solution $h_z = 0$ and $H = \lambda \cdot (\dot{y} \cos y) - h = \dot{y} \cdot h_y - h$. Therefore $H(t, x(t), u(t), \lambda(t))$ is continuous across all corners of $x(t)$, and hence it is continuous for all $t \in [t_0, t_1]$.

Corollary 3.3. If $x(t)$ is in the interior of Γ for all $t \in [t_0, t_1]$, then $\lambda(t)$ is continuous on the interval $[t_0, t_1]$.

Theorem 3.8. The Hamiltonian function evaluated at the initial and terminal points of a solution is zero. That is,

$$H(t, x(t), u(t), \lambda(t)) = 0$$

for $t = t_0$ and $t = t_1$.

Proof: We shall apply the transversality conditions of theorem 3.1.

Notice that $\phi_0 = f_0$, and recall from the proof of theorem 3.6 that

$h_z = 0$ along a solution. We also have that $h_y = \lambda \cos y$, and so

$h_y \cdot \dot{y} = (\lambda \cos y) \cdot \dot{y} = \lambda (\dot{y} \cos y) = \lambda \cdot f$ along a solution. Thus the transversality conditions yield that $\lambda_0 f_0 + \lambda \cdot f = H = 0$ for $t = t_0$ and $t = t_1$.

The next theorem follows from theorem 3.6.

Theorem 3.9. If the function $u(t)$ is differentiable except at its points of discontinuity then $H(t, x(t), u(t), \lambda(t))$ is differentiable except at these points and

$$\frac{d}{dt} H(t, x(t), u(t), \lambda(t)) = H_t(t, x(t), u(t), \lambda(t)).$$

Proof: Since $H_\lambda = f$ which is equal to $\dot{x}$ along our solution,

$$\begin{aligned} \frac{d}{dt} H &= H_t + H_x \cdot \dot{x} + H_u \cdot \dot{u} + H_\lambda \cdot \dot{\lambda} \\ &= H_t + (\dot{\lambda} + H_x) \cdot \dot{x} + H_u \cdot \dot{u} \end{aligned}$$

From theorem 3.6, on every smooth arc either $|x_i| \equiv 1$, in which case $\dot{x}_i \equiv 0$, or $\dot{\lambda}_i + H_{x_i} \equiv 0$ for $i = 1, \dots, n$, and also either $|u_j| \equiv 1$, in which case $\dot{u}_j \equiv 0$, or $H_{u_j} \equiv 0$ for $j = 1, \dots, m$. Therefore, using these facts, we see that the last two terms on the right hand side of the above equation are zero and hence, along each smooth arc of our solution $\frac{d}{dt} H = H_t$.

Corollary 3.4. In the case of an autonomous problem, that is, if the functions f_0 and f are independent of t , then

$$H(t, x(t), u(t), \lambda(t)) = 0$$

for all $t \in [t_0, t_1]$.

Proof: This follows from theorems 3.7, 3.8, and 3.9.

Theorem 3.10. Let $t \in [t_0, t_1]$ and $\gamma \in R^m$. Decompose γ into two subvectors $\bar{\gamma}$ consisting of those components of γ corresponding to coordinates of $u(t)$ for which $|u_j(t)| = 1$, and $\tilde{\gamma}$ consisting of those coordinates of $u(t)$ for which $|u_j(t)| < 1$. Decompose the vector u similarly. Then for all $\bar{\gamma}$ and for all $\tilde{\gamma}$ satisfying the equations $f_{\tilde{u}} \tilde{\gamma} = 0$,

$$\bar{\gamma} \cdot (H_{\bar{u}} * \bar{u} * \bar{\gamma}) - \tilde{\gamma} \cdot H_{\tilde{u}\tilde{u}} \tilde{\gamma} \geq 0 \text{ at time } t.$$

Proof: We shall apply the Clebsch condition of theorem 3.1. The matrices $h_{\dot{y}\dot{y}}, h_{\dot{z}\dot{y}}$ and $h_{\dot{y}\dot{z}}$ are all 0. Thus the Clebsch condition reduces to: $\rho \cdot h_{\dot{z}\dot{z}} \rho \geq 0$ for all $\rho \in R^m$ such that $\Phi_{\dot{z}} \rho = 0$. Now by differentiation and appropriate substitution we obtain $\Phi_{\dot{z}} = -f_u \text{diag}(\cos \dot{z})$ and $h_{\dot{z}\dot{z}} = \text{diag}(H_u * u) - \text{diag}(\cos \dot{z}) H_{uu} \text{diag}(\cos \dot{z})$. Therefore,

$$\begin{aligned} \rho \cdot h_{\dot{z}\dot{z}} \rho &= \rho \cdot \text{diag}(H_u * u) \rho - \rho \cdot (\text{diag}(\cos \dot{z}) H_{uu} \text{diag}(\cos \dot{z}) \rho) \\ &= \rho \cdot (H_u * u * \rho) - (\rho * \cos \dot{z}) \cdot H_{uu} (\rho * \cos \dot{z}), \quad \text{and} \\ \Phi_{\dot{z}} \rho &= -f_u (\rho * \cos \dot{z}). \end{aligned}$$

Let $\gamma \in R^m$ such that $\tilde{\gamma}$ satisfies $f_{\tilde{u}} \tilde{\gamma} = 0$. Set $\bar{\rho} = \bar{\gamma}$ and let $\tilde{\rho}$ have components $\rho_k = \frac{\gamma_k}{\cos \dot{z}_k}$. This is possible since $\cos \dot{z}_k \neq 0$ for each k such that $u_k \in \tilde{u}$. Then clearly $\tilde{\gamma} = \tilde{\rho} * \cos \dot{z}$. For each k such that $u_k \in \tilde{u}$ we have $|u_k| < 1$, and consequently by theorem 3.6, $H_{u_k}(t, x(t), u(t), \lambda(t)) = 0$. Thus, $\rho \cdot (H_u * u * \rho) = \bar{\rho} \cdot (H_{\bar{u}} * \bar{u} * \bar{\rho}) = \bar{\gamma} \cdot (H_{\bar{u}} * \bar{u} * \bar{\gamma})$ because the other terms involving components of $\tilde{\rho}$ are all zero.

Also, since $\tilde{\gamma} = \tilde{\rho} * \cos \dot{\tilde{z}}$, $f_{\tilde{u}}(\tilde{\rho} * \cos \dot{\tilde{z}}) = 0$ and since $\cos \dot{\tilde{z}}_k = 0$ for all k such that $\dot{\tilde{z}}_k \in \dot{\tilde{z}}$, we obtain $\Phi_{\dot{\tilde{z}}} \rho = -f_{\tilde{u}}(\rho * \cos \dot{\tilde{z}}) = -f_{\tilde{u}}(\tilde{\rho} * \cos \dot{\tilde{z}}) = 0$. Moreover, for the same reasons $(\rho * \cos \dot{\tilde{z}}) \cdot H_{uu}(\rho * \cos \dot{\tilde{z}}) = \tilde{\gamma} \cdot H_{\tilde{u}\tilde{u}} \tilde{\gamma}$. Therefore, by the Clebsch condition we obtain

$$\begin{aligned} \rho \cdot h_{\dot{\tilde{z}}} \rho &= \rho \cdot (H_{\tilde{u}} * u * \rho) - (\rho * \cos \dot{\tilde{z}}) \cdot H_{uu}(\rho * \cos \dot{\tilde{z}}) \\ &= \tilde{\gamma} \cdot (H_{\tilde{u}} * \tilde{u} * \tilde{\gamma}) - \tilde{\gamma} \cdot H_{\tilde{u}\tilde{u}} \tilde{\gamma} \geq 0 \end{aligned}$$

and the theorem is proved.

The next corollary is perhaps a more usable form of theorem 3.10.

Corollary 3.5. Let $t \in [t_0, t_1]$ and $\tilde{u}, \tilde{u}, \tilde{\gamma}$ and $\tilde{\gamma}$ be defined as in theorem 3.10. Then each component of the vector $H_{\tilde{u}} * \tilde{u}$ is non-negative, and for all $\tilde{\gamma}$ such that $f_{\tilde{u}} \tilde{\gamma} = 0$, $\tilde{\gamma} \cdot H_{\tilde{u}\tilde{u}} \tilde{\gamma} \leq 0$.

Proof: Successively apply theorem 3.10 with $\tilde{\gamma} = 0$ and each component of $\tilde{\gamma}$ equal to zero except for the k -th component which is set equal to one. Then $\tilde{\gamma} \cdot H_{\tilde{u}\tilde{u}} \tilde{\gamma} = 0$ and the result of theorem 3.10 reduces to $H_{\tilde{u}} u_k \geq 0$. Doing this for all k such that $u_k \in \tilde{u}$ we obtain that each component of $H_{\tilde{u}} * \tilde{u}$ is non-negative. Now apply theorem 3.10 with $\tilde{\gamma} = 0$ and $\tilde{\gamma}$ any vector such that $f_{\tilde{u}} \tilde{\gamma} = 0$. Then theorem 3.10 implies that $\tilde{\gamma} \cdot H_{\tilde{u}\tilde{u}} \tilde{\gamma} \leq 0$.

Notice that since corollary 3.5 clearly implies theorem 3.10, the condition in corollary 3.5 is actually equivalent to the condition in theorem 3.10.

Theorem 3.11. Let $t \in [t_0, t_1]$. Then for all $v \in \mathbb{R}^m$ such that $|v_i| \leq 1$ for $i = 1, \dots, m$ and $\bar{f}(t, x(t), v) = 0$ where $\bar{f}$ consists of all coordinates

of f corresponding to components of $x(t)$ for which $|x_k(t)| = 1$, the following is satisfied:

$$H(t, x(t), u(t), \lambda(t)) \geq H(t, x(t), v, \lambda(t)).$$

Proof: The Weierstrass condition in theorem 3.1 shall be applied.

Suppose $v \in \mathbb{R}^m$ such that $|v_i| \leq 1$ for $i = 1, \dots, m$ and $\bar{f}(t, x(t), v) = 0$.

Let $\dot{z} = \sin^{-1}(v)$ and $\dot{y}_k = \frac{f_k(t, x(t), v)}{\cos(y_k(t))}$ for all k such that

$|x_k(t)| < 1$ with $\dot{y}_k$ chosen arbitrarily for all k such that $|x_k(t)| = 1$.

Then $\dot{y} \cos y(t) = f(t, x(t), v) = f(t, \sin y(t), \sin \dot{z})$ since

$\cos y_k(t) = 0$ for all k such that $|x_k(t)| = 1$, and hence $(t, y(t), \dot{y}, \dot{z})$

satisfies the constraining equations. Now as was seen in the proof

of theorem 3.7, along the solution $x(t), u(t)$ we have $h_z = 0$ and

$h - \dot{y} \cdot h_y = -H$. Moreover, using the fact that $h_y = \lambda \cos y$,

$$\begin{aligned} h(t, y, \dot{y}, \dot{z}, \lambda) - \dot{y} \cdot h_y(t, y, \dot{y}, \dot{z}, \lambda) \\ &= h(t, y, \dot{y}, \dot{z}, \lambda) - (\dot{y} \cos y) \cdot \lambda \\ &= -H(t, x, v, \lambda). \end{aligned}$$

Therefore,

$E(t, y, \dot{y}, \dot{z}, \dot{y}, \dot{z}, \lambda) = H(t, x, u, \lambda) - H(t, x, v, \lambda)$ and so the Weierstrass condition yields that

$$H(t, x, u, \lambda) \geq H(t, x, v, \lambda).$$

This theorem is a generalization for the bounded state problem of the well known maximum principle of Pontryagin.

Definition 3.2. $(y(t), \dot{z}(t))$ is called a singular extremal provided

$$\frac{\partial(h_{\dot{y}}, h_{\dot{z}}, \Phi)}{\partial(\dot{y}, \dot{z}, \lambda)} \quad \left| \quad \begin{array}{l} y = y(t) \\ \dot{z} = \dot{z}(t) \end{array} \right.$$

is a singular matrix on some open interval contained in $[t_0, t_1]$.

Theorem 3.12. If $x(t)$ is on the boundary of Γ for some finite time in $[t_0, t_1]$, then $(y(t), \dot{z}(t))$ is a singular extremal.

Proof: Differentiating we obtain

$$\begin{aligned} h_{\dot{y}\dot{y}} &= \Phi_{\lambda} = 0, \quad h_{\dot{y}\dot{z}} = h_{\dot{z}\dot{y}} = 0, \\ h_{\dot{y}\lambda} &= \text{diag}(\cos y) = \Phi_{\dot{y}}, \end{aligned}$$

and

$$h_{\dot{z}\lambda} = -f_u \text{diag}(\cos \dot{z}) = \Phi_{\dot{z}}.$$

Therefore,

$$\frac{\partial(h_{\dot{y}}, h_{\dot{z}}, \Phi)}{\partial(\dot{y}, \dot{z}, \lambda)} = \begin{pmatrix} 0 & 0 & \text{diag}(\cos y) \\ 0 & h_{\dot{z}\dot{z}} & -f_u \text{diag}(\cos \dot{z}) \\ \text{diag}(\cos y) & -f_u \text{diag}(\cos \dot{z}) & 0 \end{pmatrix}$$

which is clearly singular if $\cos(y_k(t)) = 0$ on some open interval in $[t_0, t_1]$, that is, provided $x(t)$ is on the boundary of Γ for some finite time. Hence $(y(t), \dot{z}(t))$ is a singular extremal.

This theorem shows that a solution to problem II' is singular provided its corresponding solution to problem I contains a state boundary subarc. The definition of singularity given above is

equivalent to the usual definition found in the literature. The relationships between various concepts of a singular extremal is investigated by Straeter in (25).

4. EXAMPLE OF A LINEAR PROBLEM WITH BOUNDED STATE VARIABLES

The problem to be considered, a linear time-optimal control problem with bounded state variables, is the following.

Consider the differential equation $\ddot{x} = u$ where the control u is restricted by the condition $|u| \leq 1$, and the state variables x and $\dot{x}$ are restricted by $|x| \leq 1$ and $|\dot{x}| \leq 1$. To be found is a sectionally continuous function $u(t)$ which yields a solution to the above differential equation, subject to the given constraints, such that one arrives at the origin from a given initial state in the shortest possible time. That is, $u(t)$ is defined on some interval $[0, t_1]$ such that $x(0) = x_1^0$, $\dot{x}(0) = \dot{x}_2^0$ and $x(t_1) = 0$, $\dot{x}(t_1) = 0$ and t_1 is minimal.

If we let $x_1 = x$ and $x_2 = \dot{x}$, then clearly this is a special case of problem I of section 2.1 with $t_0 = 0$, $n = 2$, $m = 1$, $\Gamma = [-1, 1] \times [-1, 1]$, $\Omega = [-1, 1]$,

$$f = \begin{pmatrix} x_2 \\ u \end{pmatrix}, \quad x^0 = \begin{pmatrix} x_1^0 \\ \dot{x}_2^0 \end{pmatrix}, \quad x^1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad f_0 = 1.$$

In this chapter we shall characterize solutions to this problem by using the conditions developed in chapter 3.

First notice that in order to satisfy hypothesis (2) of theorem 3.1, we must have $|x_1(t)| < 1$ for all t according to theorem 3.2. Since f_1 is independent of u , the row in $M(t)$ corresponding to x_1 will always be zero. Thus, if $|x_1(t)| = 1$, $M(t)$ would not have

maximum rank. Moreover by theorem 3.4, the number of constraints in effect at any time cannot exceed m , which is 1 in this case, and so we can never have $|u(t)| = 1$ and $|x_2(t)| = 1$ simultaneously. In fact, if $|x_1(t)| = 1$ on some interval, then on that interval $\dot{x}_1 = 0$, in which case, by the state equations, $x_2 = 0$ on that interval, a contradiction.

Let us now investigate subarcs interior to the state region, that is, where $|x_2(t)| < 1$. We have $H = -\lambda_0 + \lambda_1 x_2 + \lambda_2 u$ and hence according to theorem 3.6, $\dot{\lambda}_1 = 0$ and $\dot{\lambda}_2 = -\lambda_1$. Moreover, by theorem 3.11, we have $\lambda_2 u(t) \geq \lambda_2 v$ for all v of absolute value less than or equal to one. Therefore, $u(t) = 1$ provided $\lambda_2(t) > 0$ and $u(t) = -1$ when $\lambda_2(t) < 0$. From the above differential equations we see that λ_1 is constant and λ_2 is linear with slope $-\lambda_1$. Notice that λ_2 is not identically zero since in this case $\lambda_1 = -\dot{\lambda}_2$ would be zero also, which would mean by corollary 3.4 that $\lambda_0 = 0$. This would violate the condition that $(\lambda_0, \lambda_1, \lambda_2) \neq 0$. Moreover, interior to the subarc, we cannot have $\lambda_2 = 0$ even at an isolated point since in this case u would have to be discontinuous at such a point, switching from -1 to $+1$ or vice versa. Hence, on such a subarc, $u(t) \equiv +1$ or $u(t) \equiv -1$. Since by theorem 3.5 $|x_2(0)| < 1$, the first subarc of a solution must be of this kind.

Type 1 ($u = 1$)

If we integrate the state equations we obtain

$$x_1 = \frac{1}{2} t^2 + c_2 t + c_1$$

$$x_2 = t + c_2$$

and elimination of t from the above yields

$$x_1 = \frac{1}{2} x_2^2 + (c_1 - \frac{1}{2} c_2^2) = \frac{1}{2} x_2^2 + c.$$

In the state plane this defines a one parameter family of parabolas with the x_1 -axis as their axis of symmetry and all opening to the right. Moreover, since $\dot{x}_2 = u = 1$, the movement along these parabolas is from bottom to top as t increases. Of course, for this problem we are only concerned with their portions lying in Γ . By applying the boundary conditions to the above we see that $c_1 = x_1^0$ and $c_2 = x_2^0$, and that it is possible to reach the origin without switching only if one starts on the parabola $x_1 = \frac{1}{2} x_2^2$ with $x_2 \leq 0$.

Type 2 ($u = -1$)

Here we obtain

$$x_1 = -\frac{1}{2} t^2 + d_2 t + d_1$$

$$x_2 = t + d_2$$

and again eliminating t yields

$$x_1 = -\frac{1}{2} x_2^2 + (d_1 + \frac{1}{2} d_2^2) = -\frac{1}{2} x_2^2 + d.$$

In the state plane this defines a one parameter family of parabolas with the x_1 -axis as their axis of symmetry and all opening to the left. Moreover, since $\dot{x}_2 = u = -1$, the phase point moves from top to bottom along one of these parabolas as t increases. Again the boundary conditions yield that $d_1 = x_1^0$ and $d_2 = x_2^0$, and it is

possible to reach the origin without switching only if one starts on the parabola $x_1 = -\frac{1}{2} x_2^2$ with $x_2 \geq 0$.

Let us now consider the possibility of subarcs on which $|x_2(t)| = 1$. In this case we must have that $\dot{x}_2 = u = 0$, according to theorem 3.6, since $|u(t)| < 1$, $\lambda_2(t) = 0$ along such a subarc, and $\dot{\lambda}_1 = 0$ as before. Integrating the state equations we see that either $x_2 = 1$ and $x_1 = t + e$ or $x_2 = -1$ and $x_1 = -t + f$ where e and f are constants. In the first case we move along the x_2 -boundary of Γ from left to right as t increases. In the second case we move from right to left.

By theorem 3.7, λ_1 must be continuous at each corner, and since it is constant on each possible type of subarc, we must have that $\lambda_1(t) = \lambda_1^0$ for all $t \in [t_0, t_1]$. Also, by corollary 3.4, $H = 0$ along the optimum trajectory. In particular this means that H is continuous across corners. Therefore, the term $\lambda_2 u$ must be continuous across corners because the other two terms in H , namely $\lambda_0 + \lambda_1 x_2$, are continuous everywhere. However, we have seen that at each corner u is discontinuous switching between its possible values of $+1$, -1 and 0 . Therefore λ_2 must be continuous at such corners with value 0 .

Let us now collect this information obtained from the necessary conditions in order to develop an optimum strategy starting at an arbitrary point for which $|x_1^0| < 1$ and $|x_2^0| < 1$. Observe that we can have at most three subarcs. We must, as was shown previously, begin on a subarc of type 1 or type 2 on which λ_2 is linear with slope $-\lambda_1$. The function λ_2 must go to 0 as we approach a switching

point. If we switch to a subarc of type 1 or type 2 again, then λ_2 can never again be 0 since on this subarc it will be linear with the same slope as previously. Thus we can never switch again. If the x_2 boundary is encountered, that is, if the absolute value of x_2 becomes 1, then we switch to $u = 0$ and λ_2 is 0 along such a subarc. In order to get to the origin we must switch from this boundary subarc to one of type 1 or type 2 in which case by the above argument λ_2 can never become 0 again. Hence no further switching is possible.

Therefore, to get to the origin from an initial point not lying on a parabola which leads to the origin we must do the following. Through (x_1^0, x_2^0) there passes exactly one member of each type 1 and type 2. However, only one of these parabolas leads (in the direction of increasing t) to a parabola which leads to the origin or to a boundary of Γ where $|x_2| = 1$ which leads to a parabola which leads to the origin. So one must travel along that parabola with $u = -1$ or $u = +1$ (whichever is appropriate) and switch the value of u to the negative of its previous value when the parabola leading to the origin is encountered. If the boundary of Γ is met, we switch to $u = 0$ and move along the boundary until arriving at the parabola leading to the origin where u is switched again to either $+1$ or -1 .

To summarize, if

$$\begin{array}{ll}
 1. \quad -1 < x_2 \leq 0 & \text{and} \left\{ \begin{array}{ll} -1 < x_1 \leq \frac{1}{2} x_2^2 & \text{use } u = +1 \\ \frac{1}{2} x_2^2 < x_1 < 1 & \text{use } u = -1 \end{array} \right. \\
 2. \quad 0 \leq x_2 < 1 & \text{and} \left\{ \begin{array}{ll} -\frac{1}{2} x_2^2 \leq x_1 < 1 & \text{use } u = -1 \\ -1 < x_1 < -\frac{1}{2} x_2^2 & \text{use } u = +1 \end{array} \right. \\
 3. \quad x_2 = 1 & \text{and} \left\{ \begin{array}{ll} -1 < x_1 < -\frac{1}{2} & \text{use } u = 0 \\ x_1 = -\frac{1}{2} & \text{use } u = -1 \end{array} \right. \\
 4. \quad x_2 = -1 & \text{and} \left\{ \begin{array}{ll} \frac{1}{2} < x_1 < 1 & \text{use } u = 0 \\ x_1 = \frac{1}{2} & \text{use } u = +1 \end{array} \right.
 \end{array}$$

However, if $x_2^0 \geq 0$ and $x_1^0 + \frac{1}{2} (x_2^0)^2 \geq 1$ the subarc of type 2 which must be taken first leads to the $x_1 = 1$ boundary which contradicts our first observation that $|x_1(t)| < 1$ at all points along a solution. Similarly, if $x_2^0 \leq 0$ and $x_1^0 - \frac{1}{2} (x_2^0)^2 \leq -1$ then the subarc of type 1 which must be used leads to the $x_1 = -1$ boundary. Hence there can be no solution to the problem for the initial point in these regions.

So we see that from each point in the interior of Γ , excluding these two regions, there emanates a unique trajectory satisfying all the necessary conditions of chapter 3 which leads to the origin. Only these trajectories can be optimal (solutions to our problem).

Thus, if a solution exists from a given initial point then it must necessarily be the above unique trajectory passing through the initial point and going to the origin. To determine if these trajectories are indeed optimal, they must be examined in the light of sufficiency criteria. Unfortunately, very little has been done in the way of developing sufficient conditions for solutions to bounded state problems. The question of sufficiency for this example with an unbounded state region is investigated in (11).

5. THE LINEAR TIME OPTIMAL CONTROL PROBLEM WITH BOUNDED STATE VARIABLES

In chapter 4 an example of a linear time optimal control problem with bounded state variables was considered. We shall take up the general problem in this chapter. That is, consider problem I in section 2.1 with

$$\begin{aligned}\Omega &= \{u \in R^m: |u_j| \leq 1 \text{ for } j = 1, 2, \dots, m\}, \\ \Gamma &= \{x \in R^n: |x_i| \leq 1 \text{ for } i = 1, 2, \dots, n\}, \\ f_0(t, x, u) &= 1 \text{ and } f(t, x, u) = Ax + Bu\end{aligned}$$

where A is an $n \times n$ matrix and B is an $n \times m$ matrix. Here the state equations are $\dot{x} = Ax + Bu$ which are linear in both the state variables and the control variables. The cost function in this case is

$$\int_{t_0}^{t_1} dt = t_1 - t_0,$$

the total time duration of a trajectory. Therefore we seek a solution to the state equation $x(t)$ and $u(t)$ going from some initial state x^0 to some final state x^1 in minimal time.

By way of notation, we shall use $N_L(A)$ to designate the left null space of a matrix. That is, $N_L(A)$ consists of all vectors λ such that $\lambda A = 0$.

5.1 Subarcs Interior to the State Region

As usual we assume that $x(t)$ and $u(t)$ defined on $[t_0, t_1]$ yield a solution to the problem. Consider a subarc of $x(t)$ from time t'

to time t'' such that $|x_i(t)| < 1$ for all $t \in (t', t'')$ and for $i = 1, \dots, n$. That is, the subarc is contained in the interior of Γ . Assume that on this subarc there are k components of $u(t)$ such that $|u_j(t)| = 1$. Let $\bar{u}(t)$ be the subvector of u consisting of all its components where $|u_j(t)| = 1$, and $\tilde{u}(t)$ consist of those components for which $|u_j(t)| < 1$. Then $\bar{u}$ will be a k -vector and $\tilde{u}$ an $(m - k)$ -vector. Let $\bar{B}$ consist of those columns of B corresponding to components of u that are in $\bar{u}$, and let $\tilde{B}$ contain the columns of B corresponding to elements of u that are in $\tilde{u}$. Then $\bar{B}$ will be an $n \times k$ matrix and $\tilde{B}$ an $n \times (m - k)$ matrix.

In chapter 2 we observed that the necessary conditions for solutions in the interior of Γ are the same as those for the problem with an unrestricted state region. Therefore, all results stated here hold also for the same problem with $\Gamma = \mathbb{R}^n$ since all results in this section follow from those necessary conditions. One of these is the well known bang - bang principle.

We shall consider λ to be a row vector so that multiplication of λ by a matrix will occur on the right. As before, x and u will be thought of as column vectors and will be multiplied on the left by matrices.

Theorem 5.1. On each such subarc $\lambda(t)$ satisfies the equations

$$\dot{\lambda} + \lambda A = 0 \text{ and } \lambda \tilde{B} = 0.$$

Proof: For this problem $H(t, x, u) = \lambda_0 + \lambda \cdot (Ax + Bu)$ and by definition $Bu = \bar{B}\bar{u} + \tilde{B}\tilde{u}$. Applying theorem 3.6, we obtain that along each subarc

of a solution $\dot{\lambda} + H_x = 0$ and $H_{u_j} = 0$ for all j such that $|u_j| < 1$ or, in other words, $H_{\tilde{u}} = 0$. Now $H_x = \lambda A$ and $H_{\tilde{u}} = \lambda \tilde{B}$. Therefore, $\dot{\lambda} + \lambda A = 0$ and $\lambda \tilde{B} = 0$.

Theorem 5.2. For all $t \in (t', t'')$, $\lambda(t) \neq 0$ and

$$\lambda(t) \in \bigcap_{i=0}^{n-1} N_L(A^{i\tilde{B}}).$$

Proof: From theorem 5.1 we have that $\lambda(t)$ satisfies $\dot{\lambda} = -\lambda A$ and $\lambda \tilde{B} = 0$. Differentiating we obtain $\dot{\lambda} \tilde{B} = 0$ and substituting for $\dot{\lambda}$ yields that $\dot{\lambda} \tilde{B} = -\lambda A \tilde{B} = 0$ along the subarc. Continuing this process we obtain $-\dot{\lambda} A \tilde{B} = \lambda A^2 \tilde{B} = 0$ along the subarc, and in general $\lambda A^{i\tilde{B}} = 0$ for $i = 0, 1, \dots, n-1$. Therefore,

$$\lambda(t) \in \bigcap_{i=0}^{n-1} N_L(A^{i\tilde{B}}).$$

If $\lambda(t) = 0$ for some t , then since by corollary 3.4 $H = 0$ along a solution, we would have at that time $0 = H = \lambda_0 + \lambda \cdot (Ax + Bu) = \lambda_0$. So $(\lambda_0, \lambda(t)) = 0$ which contradicts the result of theorem 3.6 that $(\lambda_0, \lambda(t)) \neq 0$ for all t . Therefore, $\lambda(t) \neq 0$ for all $t \in (t', t'')$.

Corollary 5.1. The rank of the $n \times n(m-k)$ matrix $[\tilde{B}, A\tilde{B}, \dots, A^{n-1}\tilde{B}]$ is less than n .

Proof: By theorem 5.2, $\lambda \in \bigcap_{i=0}^{n-1} N_L(A^{i\tilde{B}})$ on the subarc. That is,

$$\lambda A^{i\tilde{B}} = 0 \text{ for } i = 0, 1, \dots, n-1 \text{ or equivalently } \lambda[\tilde{B}, A\tilde{B}, \dots, A^{n-1}\tilde{B}] = 0.$$

But since $\lambda \neq 0$ this means that the rank of the above matrix must be less than n .

By using this corollary we can determine which combinations of

components of u can possibly have absolute value one on a subarc.

If, for a given choice of $\tilde{B}$, the rank of the above matrix is n , then we can never have a solution to the problem with such a $\bar{u}$ and $\tilde{u}$ on any subarc.

Definition 5.1. The matrices A and B satisfy the general position condition of order q where $1 \leq q \leq m$ provided whenever Q is an $n \times q$ matrix consisting of q columns of the matrix B , then the $n \times nq$ matrix $[Q, AQ, A^2Q, \dots, A^{n-1}Q]$ has rank n . If A and B satisfy the general position condition of order 1, then we simply say that they satisfy the general position condition.

Theorem 5.3. A and B satisfy the general position condition of order q if and only if for all matrices Q consisting of q columns of B , we have

$$\bigcap_{i=0}^{n-1} N_L(A^i Q) = \{0\}$$

Proof: Suppose $\alpha \in \bigcap_{i=0}^{n-1} N_L(A^i Q)$ for some Q . Then $\alpha A^i Q = 0$ for $i = 0, \dots, n-1$, and hence

$$\alpha [Q, AQ, \dots, A^{n-1}Q] = 0$$

which implies that $\alpha = 0$ by the general position condition of order q .

Therefore $\bigcap_{i=0}^{n-1} N_L(A^i Q) = \{0\}$. If the general position condition of order q is not satisfied, then there exists $\alpha \neq 0$ such that for some Q , $\alpha [Q, AQ, \dots, A^{n-1}Q] = 0$ which implies that $\alpha A^i Q = 0$ for $i = 0, \dots, n-1$.

But this means that $\alpha \in \bigcap_{i=0}^{n-1} N_L(A^i Q) = \{0\}$, a contradiction. Hence

the general position condition of order q is satisfied.

Theorem 5.4. If A and B satisfy the general position condition of order q , then A and B satisfy the general position condition of order q' for all q' such that $q \leq q' \leq m$.

Proof: Let Q' be an $n \times q'$ matrix consisting of q' columns of B . Then since $q \leq q'$, let Q be the $n \times q$ matrix consisting of the first q columns of Q' . By the hypothesis the matrix $[Q, AQ, \dots, A^{n-1}Q]$ has rank n . Therefore, $[Q', AQ', \dots, A^{n-1}Q']$ has rank n also since the previous matrix consists of a subset of the columns of this matrix. Hence A and B satisfy the general position condition of order q' .

Theorem 5.5. Suppose A and B satisfy the general position condition of order q . Then along any subarc of a solution interior to Γ , at least $m - q + 1$ controls have absolute value one. That is,

$$k \geq m - q + 1.$$

Proof: We shall argue by contradiction. Suppose $k \leq m - q$, then $q \leq m - k$, in which case by theorem 5.4, we have that A and B satisfy the general position condition of order $m - k$. But $\tilde{B}$ consists of exactly $m - k$ columns of B , which means by corollary 5.1 that

$$\text{rank} ([\tilde{B}, A\tilde{B}, \dots, A^{n-1}\tilde{B}]) < n,$$

a contradiction to the fact that A and B satisfy the general position condition of order $m - k$. Therefore, $k \geq m - q + 1$.

Corollary 5.2. (bang - bang principle) If A and B satisfy the general position condition, then along any subarc of a solution interior to Γ we have that $|u_j| = 1$ for $j = 1, \dots, m$.

Proof: By theorem 5.5, using $q = 1$, we obtain $k \geq m - 1 + 1 = m$, and since k can be at most m , we thus have $k = m$. Therefore, all components of the control vector have absolute value one.

Corollary 5.3. If A and B satisfy the general position condition of order q where $1 \leq q \leq m$, then the controls along any subarc interior to Γ always lie on the boundary of Ω .

Proof: By theorem 5.5, $k \geq m - q + 1 \geq m - m + 1 = 1$. Therefore at least one component of u has absolute value one. Hence u lies on the boundary of Ω .

(Note: if $q = m$, then the hypothesis of corollary 5.3 is the so-called controllability condition.)

Corollary 5.4. If there is a subarc of the optimal trajectory on which the control is interior to Ω , then A and B do not satisfy the general position condition of any order.

We shall now develop a method for determining which components of u must have absolute value one on every subarc of $x(t)$ which is interior to Γ .

Designate the columns of B by $b_1, b_2, \dots, b_m$ and define

$$M_j = [b_j, Ab_j, A^2b_j, \dots, A^{n-1}b_j]$$

for $j = 1, \dots, m$. Clearly each M_j is an $n \times n$ matrix.

Notice that:

- (a) A and B satisfy the general position condition if and only if each M_j has rank n .
- (b) A and B satisfy the general position condition of order q if and only if each $n \times nq$ matrix consisting of precisely q of the M_j 's has rank n . This follows because if $Q = [b_{j_1}, \dots, b_{j_q}]$, then $[Q, AQ, \dots, A^{n-1}Q]$ is just $[M_{j_1}, \dots, M_{j_q}]$ with its columns rearranged. Therefore both matrices have the same rank.

Theorem 5.6. Suppose M_j has rank n for some j where $1 \leq j \leq m$, then $|u_j| = 1$ on any optimal subarc which is interior to Γ .

Proof: Suppose $|u_j| < 1$ on some subarc interior to Γ . Then on that subarc from theorem 5.1 we have

$$\lambda \in \bigcap_{i=0}^{n-1} N_L(A^i \tilde{B})$$

where $\tilde{B}$ contains column b_j since $|u_j| < 1$. Since M_j has rank n , it follows that $[\tilde{B}, A\tilde{B}, \dots, A^{n-1}\tilde{B}]$ has rank n because M_j consists of a subset of the columns of this matrix. This was shown in (b) above.

Therefore $\bigcap_{i=0}^{n-1} N_L(A^i \tilde{B}) = \{0\}$, which means that $\lambda = 0$ on this subarc, a contradiction. Thus we must have that $|u_j| = 1$.

Corollary 5.5. If $|u_j| < 1$ on some subarc interior to Γ , then M_j has rank less than n .

Theorem 5.7. On each subarc of $x(t)$ interior to Γ ,

$$u_j = \begin{cases} -1 & \text{provided } \lambda b_j < 0 \\ +1 & \text{provided } \lambda b_j > 0 \end{cases}$$

for $j = 1, \dots, m$.

Proof: According to theorem 5.1, $\lambda b_j = 0$ for all j such that u_j is in $\tilde{u}$. Thus, we need only investigate the u_j that are in $\bar{u}$, that is, the ones that have absolute value one. According to corollary 3.5, each component of $H_{\bar{u}} \bar{u}$ must be non-negative. But $H_{\bar{u}} = \lambda \bar{B}$; therefore this condition requires that $\lambda b_j u_j \geq 0$ for each u_j in $\bar{u}$. Hence when $\lambda b_j > 0$, $u_j = +1$, and when $\lambda b_j < 0$, $u_j = -1$.

5.2 Subarcs Along the Boundary of the State Region

We shall now consider a subarc of $x(t)$ from time t' to time t'' such that $|x_i(t)| = 1$ for some i such that $1 \leq i \leq n$ and for all $t \in (t', t'')$. That is, the subarc is contained in the boundary of Γ . Assume that along this subarc there are ℓ components of x such that $|x_i| = 1$. Let $\bar{x}$ be the subvector of x consisting of all x_i such that $|x_i| = 1$, and $\tilde{x}$ contain all x_i such that $|x_i| < 1$ on the subarc. Then $\bar{x}$ is an ℓ -vector and $\tilde{x}$ an $(n - \ell)$ -vector. Define $\bar{u}$ and $\tilde{u}$ as in section 5.1.

Now let $\bar{A}$ consist of all columns of A corresponding to components of x in $\bar{x}$, and $\tilde{A}$ contain all columns of A corresponding to components of x in $\tilde{x}$. Subdivide $\bar{A}$ into A_1 and A_3 , A_1 consisting of all of its rows corresponding to the x_i in $\bar{x}$ and A_3 containing all rows

corresponding to the x_i in $\tilde{x}$. Similarly, subdivide $\tilde{A}$ into A_2 and A_4 .

Then: $\bar{A}$ is an $n \times \ell$ matrix;
 $\tilde{A}$ is an $n \times (n - \ell)$ matrix;
 A_1 is an $\ell \times \ell$ matrix;
 A_2 is an $\ell \times (n - \ell)$ matrix;
 A_3 is an $(n - \ell) \times \ell$ matrix;
 A_4 is an $(n - \ell) \times (n - \ell)$ matrix.

Define $\bar{B}$ and $\tilde{B}$ as in section 5.1, and subdivide $\bar{B}$ into B_1 and B_3 . Let B_1 consist of all rows of $\bar{B}$ corresponding to components of x in $\bar{x}$, and B_3 contain all rows corresponding to the x_i in $\tilde{x}$. Subdivide $\tilde{B}$ into B_2 and B_4 similarly.

Then: $\bar{B}$ is an $n \times k$ matrix;
 $\tilde{B}$ is an $n \times (m - k)$ matrix;
 B_1 is an $\ell \times k$ matrix;
 B_2 is an $\ell \times (m - k)$ matrix;
 B_3 is an $(n - \ell) \times k$ matrix;
 B_4 is an $(n - \ell) \times (m - k)$ matrix.

Theorem 5.8. On any subarc along the boundary of Γ , B_2 has rank ℓ .

Proof: By theorem 3.2, along this subarc $M(t)$ must have rank ℓ .

Recall that $\bar{M}(t)$ consists of the rows of $\bar{f}_u \text{diag}(\cos \dot{z})$ corresponding to x_i for which $|x_i| = 1$. Therefore $M(t) = \bar{f}_u \text{diag}(\cos \dot{z})$. For all u_j in $\bar{u}$, $\cos \dot{z}_j = 0$. Thus $\text{rank}(M(t)) = \text{rank}(\bar{f}_u \text{diag}(\cos \dot{z}))$ since the j -th column in $\bar{f}_u$ is multiplied by $\cos \dot{z}_j$. Also, for all

u_j in $\tilde{u}$, $\cos \dot{z}_j \neq 0$, so by dividing each column of $\bar{f}_{\tilde{u}} \text{diag} (\cos \dot{z})$ by the appropriate $\cos \dot{z}_j \neq 0$, we obtain

$$\text{rank } (\bar{M}(t)) = \text{rank } (\bar{f}_{\tilde{u}} \text{diag} (\cos \dot{z})) = \text{rank } (\bar{f}_{\tilde{u}}).$$

Now $f = Ax + Bu = Ax + \bar{B}\bar{u} + \tilde{B}\tilde{u}$, so $f_{\tilde{u}} = \tilde{B}$. Eliminating all rows corresponding to x_i in $\bar{x}$, we obtain $\bar{f}_{\tilde{u}} = B_2$. Therefore,
 $\ell = \text{rank } (\bar{M}(t)) = \text{rank } (B_2)$.

Theorem 5.9. On each subarc lying on the boundary of Γ , $x(t), u(t)$ and $\lambda(t)$ satisfy the equations

$$\begin{aligned} \bar{\lambda} B_2 + \tilde{\lambda} B_4 &= 0 \\ \dot{\tilde{\lambda}} + \bar{\lambda} A_2 + \tilde{\lambda} A_4 &= 0, \\ A_1 \bar{x} + A_2 \tilde{x} + B_1 \bar{u} + B_2 \tilde{u} &= 0, \\ A_3 \bar{x} + A_4 \tilde{x} + B_3 \bar{u} + B_4 \tilde{u} &= \dot{\tilde{x}}, \\ \lambda_0 + \tilde{\lambda} (A_3 \bar{x} + A_4 \tilde{x} + B_3 \bar{u} + B_4 \tilde{u}) &= 0. \end{aligned}$$

The last equation holds if u is differentiable on the subarc.

Proof: For this problem

$$\begin{aligned} H &= \lambda_0 + \lambda \cdot (Ax + Bu) \\ &= \lambda_0 + \lambda \cdot (\bar{A}\bar{x} + \tilde{A}\tilde{x} + \bar{B}\bar{u} + \tilde{B}\tilde{u}) \\ &= \lambda_0 + \bar{\lambda} A_1 \bar{x} + \tilde{\lambda} A_3 \bar{x} + \bar{\lambda} A_2 \tilde{x} + \tilde{\lambda} A_4 \tilde{x} \\ &\quad + \bar{\lambda} B_1 \bar{u} + \tilde{\lambda} B_3 \bar{u} + \bar{\lambda} B_2 \tilde{u} + \tilde{\lambda} B_4 \tilde{u}. \end{aligned}$$

Hence, applying theorem 3.6,

$$\begin{aligned} H_{\tilde{u}} &= \bar{\lambda} B_2 + \tilde{\lambda} B_4 = 0 \quad \text{and} \\ \dot{\tilde{\lambda}} + H_{\tilde{x}} &= \dot{\tilde{\lambda}} + \bar{\lambda} A_2 + \tilde{\lambda} A_4 = 0. \end{aligned}$$

So, the first two equations are proved. For each x_i in $\bar{x}$, $\dot{x}_i = 0$ since x_i is constant along the subarc. Thus, $\bar{f} = 0$ along the subarc. But

$$f = Ax + Bu = \bar{A}x + \tilde{A}\tilde{x} + \bar{B}u + \tilde{B}\tilde{u},$$

and hence $\bar{f} = A_1\bar{x} + A_2\tilde{x} + B_1\bar{u} + B_2\tilde{u} = 0$. For the other components $\dot{\tilde{x}} = \tilde{f} = A_3\bar{x} + A_4\tilde{x} + B_3\bar{u} + B_4\tilde{u}$ are the state equations. If u is differentiable on the subarc, then we can apply corollary 3.4 to obtain $H = 0$ along the subarc. Now by subtracting from the expansion of H above the terms in the third equation, we obtain

$$\begin{aligned} H &= \lambda_0 + \tilde{\lambda}A_3\bar{x} + \tilde{\lambda}A_4\tilde{x} + \tilde{\lambda}B_3\bar{u} + \tilde{\lambda}B_4\tilde{u} \\ &= \lambda_0 + \tilde{\lambda}(A_3\bar{x} + A_4\tilde{x} + B_3\bar{u} + B_4\tilde{u}) = 0. \end{aligned}$$

Theorem 5.10. Let $\tilde{\lambda}' = \tilde{\lambda}(t')$, then for all $t \in (t', t'')$, $\tilde{\lambda}(t) \neq 0$ and

$$\begin{aligned} \tilde{\lambda}(t) &= \tilde{\lambda}' e^{(B_4B_2^TKA_2 - A_4)(t - t')}, \\ \bar{\lambda}(t) &= -\tilde{\lambda}(t)B_4B_2^TK, \\ \tilde{\lambda}(t) &\in N_L(B_4 - B_4B_2^TKB_2), \end{aligned}$$

where $K = (B_2B_2^T)^{-1}$.

Proof: Since B_2 has rank ℓ , $B_2B_2^T$ is an invertible $\ell \times \ell$ matrix.

Thus, K is well defined. By theorem 5.8, $\bar{\lambda}B_2 = -\tilde{\lambda}B_4$ and so

$\bar{\lambda}B_2B_2^T = -\tilde{\lambda}B_4B_2^T$ which implies that $\bar{\lambda} = -\tilde{\lambda}B_4B_2^TK$. Also, by theorem

5.8, $\dot{\tilde{\lambda}} = -\bar{\lambda}A_2 - \tilde{\lambda}A_4$. Therefore, substituting for $\bar{\lambda}$, we obtain

$\dot{\tilde{\lambda}} = \tilde{\lambda}(B_4B_2^TKA_2 - A_4)$ which has the unique solution

$$\tilde{\lambda}(t) = \tilde{\lambda}' e^{(B_4 B_2^T K A_2 - A_4)(t - t')}$$

on the subarc. Also, if $\tilde{\lambda}(t) = 0$ for some $t \in (t', t'')$, then by the second equation of the theorem, $\bar{\lambda}(t) = 0$. Moreover, by viewing the fifth equation of theorem 5.8, we see that $\lambda_0 = 0$. Hence $(\lambda_0, \lambda) = 0$ which violates the conclusion of theorem 3.6 that $(\lambda_0, \lambda) \neq 0$ for all $t \in [t_0, t_1]$. Thus $\tilde{\lambda}(t) \neq 0$ for all $t \in (t', t'')$. Finally if we take $\bar{\lambda} = -\tilde{\lambda} B_4 B_2^T K$ and substitute it into $\bar{\lambda} B_2 + \tilde{\lambda} B_4 = 0$, we obtain

$$\tilde{\lambda}(B_4 - B_4 B_2^T K B_2) = 0.$$

It thus follows that $\tilde{\lambda} \in N_L(B_4 - B_4 B_2^T K B_2)$.

Corollary 5.6. If $k + \ell = m$, then $\tilde{\lambda}(t) \in N_L(B_4 - B_4 B_2^T K B_2)$ is always satisfied.

Proof: B_2 is an $\ell \times (m - k)$ matrix. Therefore, if $\ell = m - k$, B_2 is a square $\ell \times \ell$ matrix, and since it has rank ℓ , it is invertible and so is B_2^T . Hence, $K = (B_2 B_2^T)^{-1} = (B_2^T)^{-1} B_2^{-1}$. Thus $B_2^T K B_2 = B_2^T (B_2^T)^{-1} B_2^{-1} B_2 = I$, the $\ell \times \ell$ identity matrix. It therefore follows that

$$B_4 - B_4 B_2^T K B_2 = B_4 - B_4 = 0,$$

the $(n - \ell) \times \ell$ zero matrix, and so

$$\tilde{\lambda}(t) \in N_L(0) = R^{n-\ell}$$

is automatically satisfied for all $t \in (t', t'')$.

Notice that in case $k + \ell < m$, the above condition is not necessarily trivial. In this case $\ell < m - k$ which implies that $B_2^T K B_2$ has rank less than $m - k$ since B_2 has rank ℓ . Hence $B_2^T K B_2$ cannot be the $(m - k) \times (m - k)$ identity matrix.

Corollary 5.7. The rank of $B_4 - B_4 B_2^T K B_2$ is less than $n - \ell$.

Proof: This follows from the two results of theorem 5.9 that $\tilde{\lambda}(t) \neq 0$ and $\tilde{\lambda}(t) \in N_L(B_4 - B_4 B_2^T K B_2)$.

Theorem 5.11. Let $C = \tilde{A} B_4 B_2^T K$, an $n \times \ell$ matrix, and form the $n \times n$ matrix D in the following way. For each i such that $1 \leq i \leq n$, if $|x_i| = 1$ and x_i is the j -th component of $\bar{x}$, then let the i -th column of D be the j -th column of C . If $|x_i| < 1$ and x_i is the j -th component of $\tilde{x}$, then let the i -th column of D be the j -th column of $-\tilde{A}$. Then along this subarc, $\lambda(t)$ satisfies

$$\dot{\lambda} = \lambda D \text{ and } \lambda \tilde{B} = 0.$$

Proof: By definition $\bar{\lambda} B_2 + \tilde{\lambda} B_4 = \lambda \tilde{B}$, so the second equation is satisfied by theorem 5.9. Moreover, if we differentiate $\bar{\lambda} B_2 + \tilde{\lambda} B_4 = 0$, we obtain

$$\dot{\bar{\lambda}} B_2 = -\dot{\tilde{\lambda}} B_4 = (\bar{\lambda} A_2 + \tilde{\lambda} A_4) B_4 \text{ since } \dot{\lambda} = -\bar{\lambda} A_2 - \tilde{\lambda} A_4.$$

Therefore, $\dot{\bar{\lambda}} = (\bar{\lambda} A_2 + \tilde{\lambda} A_4) B_4 B_2^T K$ which is equal to $\lambda \tilde{A} B_4 B_2^T K$ by the definition of A_2 and A_4 . So we have $\dot{\bar{\lambda}} = \lambda C$ and $\dot{\tilde{\lambda}} = \lambda(-\tilde{A})$, and putting these together according to the definition of D in the hypothesis, it follows that $\dot{\lambda} = \lambda D$.

We have now put the equations for λ on a subarc along the boundary of the state region into the form of theorem 5.1, which gives the equation for λ on a subarc interior to the state region. Thus, the same type of logic used in theorem 5.2 may now be applied to obtain similar results in this case.

Theorem 5.12. For all $t \in (t', t'')$, $\lambda(t) \in \bigcap_{i=0}^{n-1} N_L(D^i \tilde{B})$.

Proof: By theorem 5.10, $\lambda(t)$ satisfies the equations $\dot{\lambda} = \lambda D$ and $\lambda \tilde{B} = 0$ on this subarc. Thus, by differentiating we obtain $\dot{\lambda} \tilde{B} = 0$ and since $\dot{\lambda} = \lambda D$, this implies that $\lambda D \tilde{B} = 0$. Continuing this process, we obtain $\dot{\lambda} D \tilde{B} = \lambda D^2 \tilde{B} = 0$ along the subarc, and in general $\lambda D^i \tilde{B} = 0$ for $i = 0, 1, \dots, n-1$. Therefore,

$$\lambda(t) \in \bigcap_{i=0}^{n-1} N_L(D^i \tilde{B}).$$

Corollary 5.8. The rank of the $n \times n(m-k)$ matrix $[\tilde{B}, D\tilde{B}, \dots, D^{n-1}\tilde{B}]$ is less than n .

Proof: This follows from the facts that $\lambda(t) \in \bigcap_{i=0}^{n-1} N_L(D^i \tilde{B})$, and $\lambda(t) \neq 0$.

Notice that there is a one-to-one correspondence between possible boundary configurations and decompositions of A and B into $A_1, A_2, A_3, A_4, B_1, B_2, B_3$, and B_4 . By possible boundary configurations we mean the possible combinations of each component of u and x either having absolute value equal to one or less than one. The theory of this section was developed with the thought in mind of enabling one to narrow the list of possible boundary configurations somewhat.

First one applies the condition that $k + \ell \leq m$ to eliminate all cases where the total number of constraints in effect exceeds the number of control variables. One might observe that this is the same as saying that the total number of constraints not in effect must be greater than or equal to the number of state variables. This is true because if $k + \ell \leq m$ then $-k - \ell \geq -m$, and if we add $n + m$ to both sides of this inequality we obtain $(m - k) + (n - \ell) \geq n$.

Secondly, one can apply corollary 5.7 to rule out some or all of the combinations where $k + \ell < m$. This is a consistency condition which must be satisfied if there is to be a non-trivial solution to the equation for λ in theorem 5.8.

Finally, one can reduce the possibilities further by applying corollary 5.7.

Theorem 5.13. For all $t \in [t', t'']$ and for all $v \in \Omega$ such that $A_1 \bar{x}(t) + A_2 \tilde{x}(t) + B_1 \bar{v} + B_2 \tilde{v} = 0$, we have

$$\tilde{\lambda}(t)(B_3 \bar{u}(t) + B_4 \tilde{u}(t)) \geq \tilde{\lambda}(t)(B_3 \bar{v} + B_4 \tilde{v}).$$

Proof: From the proof of theorem 5.9, we have that if

$$A_1 \bar{x}(t) + A_2 \tilde{x}(t) + B_1 \bar{v} + B_2 \tilde{v} = 0,$$

then

$$H(x(t), v, \lambda(t)) = \lambda_0 + \tilde{\lambda}(t)(A_3 \bar{x}(t) + A_4 \tilde{x}(t) + B_3 \bar{v} + B_4 \tilde{v}).$$

Thus, applying theorem 3.11, it follows that

$$\lambda_0 + \tilde{\lambda}(t)(A_3 \bar{x}(t) + A_4 \tilde{x}(t) + B_3 \bar{u}(t) + B_4 \tilde{u}(t)) \geq \\ \lambda_0 + \tilde{\lambda}(t)(A_3 \bar{x}(t) + A_4 \tilde{x}(t) + B_3 \bar{v} + B_4 \tilde{v}),$$

and therefore, by subtracting out the common terms on each side of the inequality, we obtain

$$\tilde{\lambda}(t)(B_3 \bar{u}(t) + B_4 \tilde{u}(t)) \geq \tilde{\lambda}(t)(B_3 \bar{v} + B_4 \tilde{v}).$$

Theorem 5.14. On each subarc of $x(t)$ along the boundary of Γ ,

$$u_j = \begin{cases} -1 & \text{provided } \lambda b_j < 0 \\ +1 & \text{provided } \lambda b_j > 0 \end{cases}$$

for $j = 1, \dots, m$ where b_j is in the j -th column of the matrix B .

Proof: According to theorem 5.11, $\lambda b_j = 0$ for all j such that u_j is in $\tilde{u}$. Thus, we need only investigate the u_j that are in $\bar{u}$, that is, the ones that have absolute value one. According to corollary 3.5, each component of $H_{\bar{u}} \bar{u}$ must be non-negative. But $H_{\bar{u}} = \lambda \bar{B}$ and therefore this condition requires that $\lambda b_j u_j \geq 0$ for each u_j in $\bar{u}$. Hence, when $\lambda b_j > 0$ then $u_j = +1$ and when $\lambda b_j < 0$, $u_j = -1$.

So we see that the results of theorem 5.7 hold even in the case where $x(t)$ lies on the boundary of Γ .

6. SUMMARY AND CONCLUSIONS

In this thesis a method has been described whereby an optimal control problem with bounded state variables may be transformed into an equivalent Lagrange problem. This was accomplished by means of differentiable mappings which take some Euclidean space onto the closed and bounded control and state regions. Whereas all such mappings lead to a Lagrange problem, it has been shown that only those which were defined as acceptable pairs of transformations are suitable in the sense that solutions to the Lagrange problem lead to solutions to the bounded state problem and vice versa. In particular, an acceptable pair of transformations was exhibited for the case when the control and state regions are right parallelepipeds.

The necessary conditions of the Calculus of Variations were then applied to the transform of the non-linear, non-autonomous optimal control problem with the unit n -cube and unit m -cube as state and control regions respectively. These conditions are the Euler Lagrange equations, corner conditions, transversality conditions, the Clebsch condition and the Weierstrass condition. In each case analogous conditions were developed which solutions of the bounded state problem must satisfy. In addition, the hypothesis for applying these conditions led to theorems which state that the initial and terminal points must be interior to the state region, and that a certain submatrix of the matrix of partials of the constraining

differential equations with respect to the control variables must have maximum row rank.

These results were then applied to a simple example of a linear time optimal control problem with bounded state variables for which solutions were completely characterized. The general linear time optimal problem was then considered and various necessary conditions developed for it in terms of the coefficient matrices.

First, the properties of subarcs of a solution interior to the state region were developed. It was observed that these must also hold for the non-state constrained problem. Among them is the well known bang - bang principle. Then subarcs which lie along the boundary of the state region were investigated with several new necessary conditions developed.

There are a number of areas in which the author feels that the research of this thesis may be extended to yield further results. If acceptable pairs of transformations can be exhibited for more general state and control regions, perhaps convex polyhedra or even compact convex sets, then the same analysis can be applied to obtain necessary conditions. Of particular interest would be state regions described by a system of inequality constraints of the form $g(x) \leq 0$.

The question of sufficiency has been totally ignored in this thesis. Although there are a number of very difficult problems associated with the development of sufficient conditions for the bounded state problem, the author feels that some results may be

obtained by using this transformation method to apply the sufficient conditions of the Calculus of Variations to the bounded state problem. The question of existence of solutions may possibly be studied by applying the theory of generalized curves to the transformed problem.

It was shown in chapter 3 that state boundary subarcs are singular subarcs of the transformed Lagrange problem. Therefore, an investigation of the known properties of singular subarcs may lead to new necessary conditions for the bounded state problem.

It was observed that a number of the results of chapter 3 are similar to those obtained by other authors. Further studies should be made to determine the exact relationship between the results contained herein and those of others, notably Berkovitz and Gamkrelidze.

In this formulation, sectionally continuous controls and sectionally smooth trajectories were considered. The same theory could have been developed using bounded measurable controls and absolutely continuous trajectories. Also, other types of problems could be investigated. Perhaps the best place to start would be to consider the linear problem with different cost functions, for example, a quadratic cost function.

Finally, some experimentation should be made, possibly on a computer, in the utilization of the conditions developed to obtain solutions to particular problems.

7. LIST OF REFERENCES

1. Berkovitz, L. D. 1961. Variational Methods in Problems of Control and Programming. *Journal of Mathematical Analysis and Applications* 3: 145-169.
2. Berkovitz, L. D. 1962. On Control Problems with Bounded State Variables. *Journal of Mathematical Analysis and Applications* 5: 488-498.
3. Berkovitz, L. D. and Dreyfus, S. E. 1965. The Equivalence of Some Necessary Conditions for Optimal Control in Problems with Bounded State Variables. *Journal of Mathematical Analysis and Applications* 10: 275-283.
4. Bryson, A. E., Denham, W. F. and Dreyfus, S. E. 1963. Optimal Programming Problems with Inequality Constraints I: Necessary Conditions for Extremal Solutions. *AIAA Journal* 1 (11): 2544-2550.
5. Caratheodory, C. 1967. *Calculus of Variations and Partial Differential Equations of the First Order, Volume II.* Holden Day, Inc., San Francisco, California.
6. Chang, S. S. L. 1962. Optimal Control in Bounded Phase Space. *Automatica* 1: 55-67.
7. Chang, S. S. L. 1962. An Extension of Ascoli's Theorem and Its Application to the Theory of Optimal Control. AROSR Report 1973.
8. Dreyfus, S. E. 1962. Variational Problems with State Variable Inequality Constraints. Ph.D. Dissertation, Harvard University.
9. Dreyfus, S. E. 1962. Variational Problems with Inequality Constraints. *Journal of Mathematical Analysis and Applications* 4: 297-308.
10. Guinn, T. 1965. The Problem of Bounded Space Coordinates as a Problem of Hestenes. *SIAM J. Control* 3 (2): 181-190.
11. Hanafy, L. M. 1970. The Linear Time Optimal Control Problem from a Calculus of Variations Point of View. NASA CR - 1612.

12. Hestenes, M. R. 1965. On Variational Theory and Optimal Control Theory. SIAM J. Control 3 (1): 23-48.
13. Hestenes, M. R. 1966. Calculus of Variations and Optimal Control Theory. John Wiley and Sons, Inc., New York.
14. Jacobson, D. H., Lele, M. M. and Speyer, J. L. 1969. New Necessary Conditions of Optimality for Control Problems with State-Variable Inequality Constraints. Harvard University, Division of Engineering and Applied Physics Technical Rept. No. 597.
15. Kalman, R. E. 1961. The Theory of Optimal Control and the Calculus of Variations, RIAS Tech. Rept. 61-3, Research Institute for Advanced Studies, Baltimore, Maryland.
16. Khrustalev, M. M. 1967. Sufficient Conditions for Optimality in Problems with Constraints on Phase Coordinates. Advtomatika i Telemekhanika 4: 18-29.
17. McIntire, J. and Paiewonsky, B. 1964. On Optimal Control with Bounded State Variables. Aeronautical Research Associates of Princeton Rept. 60.
18. Park, S. K. 1970. On the Equivalence of Optimal Control Problems and the Transformation of Optimal Control Problems with Compact Control Regions into Lagrange Problems. Ph.D. Thesis, North Carolina State University.
19. Park, S. K. 1970. Optimal Control Problems as Equivalent Lagrange Problems. NASA CR - 1550.
20. Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V. and Mishchenko, E. F. 1962. The Mathematical Theory of Optimal Processes. Interscience Publishers, New York.
21. Russak, I. B. 1970. On Problems with Bounded State Variables. Journal of Optimization Theory and Applications 5 (2): 114-157.
22. Russell, D. L. 1965. Penalty Functions and Bounded Phase Coordinate Control. SIAM J. Control 2 (3): 409-422.
23. Sagan, Hans. 1969. Introduction to the Calculus of Variations. McGraw Hill Book Co., New York.

24. Speyer, J. L. and Bryson, A. E. 1968. Optimal Programming Problems with a Bounded State Space 6 (8): 1488-1491.
25. Straeter, T. A. 1970. Singular Extremaloids in Optimal Control Theory and the Calculus of Variations. NASA CR - 1581.
26. Valentine, F. A. 1937. The Problem of Lagrange with Differential Inequalities as Side Conditions. Contributions to the Calculus of Variations 1933-37, University of Chicago Press, Chicago, Illinois.