AN ABSTRACT OF THE THESIS OF

Weldon A. Lodwick for the degree of Doctor of Philosophy

in Mathematics presented on September 14, 1979

Title: Two Numerical Methods for the Solutions of Optimal Control Problems With Computed Error Bounds Using the Maximum Principle of Pontryagin

Signature redacted for privacy.

Abstract approved: ______________________

Joel Davis

A simplification of the proof of the maximum principle of Pontryagin is obtained for constrained and unconstrained optimal control problems. Two numerical methods for solving optimal control problems with guaranteed error bounds using the maximum principle of Pontryagin and interval analysis are derived.

Two Numerical Methods for the Solutions of Optimal Control Problems With Computed Error Bounds Using the Maximum Principle of Pontryagin

by

Weldon A. Lodwick

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Commencement June 1980

APPROVED:

Signature redacted for privacy.

Associate Professor of Mathematics

in charge of major

Signature redacted for privacy.

Chairperson of Mathematics

Signature redacted for privacy.

Dean of Graduate School

Date thesis is presented September 14, 1979

Typed by Aimée Alysa Harris for Weldon A. Lodwick

ACKNOWLEDGEMENTS

First thanks must go to my father and mother who showed me a love for others which transcends cultural bounds and whose daily victory over suffering is an example for my life. I thank Dora Grady, my wife, for being a friend, lover, and proud woman. To my son I owe many thanks for sharing his joy of life with me. I also thank Uncle Walter and Aunt Mary who not only stimulated me intellectually, but showed me that love for one another is love for others.

I want to thank my friends who shared in my life at Oregon State University and I quote a Brasilian poet: "Marks in the sand are taken away by the wind and the sea, but your marks upon my soul are forever with me." I thank Vanderlei Canhos, Martin Stynes, Bob Clark, Brian McCay, S.P.S. Bhatia, Jeanne Ng, Kate Reedy, Gary Greenfield, Dave Vomocil, John Holt, and Tom Kelly.

I thank Ed Cracraft, Barbi Donnell, Jolan Eross, and Donna Templeton, whom I consider friends. Their service to each of us in the mathematics department has been too long taken for granted.

I thank Dr. S.P.S. Bhatia and Dr. Peter Frick for introducing me to optimal control theory. I especially thank Martin Stynes for his encouragement during the most difficult hours of my thesis writing.

I thank Joel Davis for giving me the freedom to investigate optimal control theory, and tackling the often thankless job of being critic, advisor, and major professor.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd.)

Two Numerical Methods for the Solutions of Optimal Control Problems With Computed Error Bounds Using the Maximum Principle of Pontryagin

0. INTRODUCTION

0.1 Optimal Control

Optimal control is a dynamic optimization problem. The optimal control approach to dynamic optimization divides the problem into states and controls. At each instant of time the controls have an effect on the change with respect to time of the states of the system. The system is characterized by the states at each instant in time. The system has a performance measure called the cost functional and it is this cost functional which measures and determines the optimal condition for the system.

The controls can be restricted to belong to a subset of its defining space in which case the control problem is constrained. Otherwise, it is unconstrained. The type of differential equation that describes the states as well as the type of functional which describes the performance measure will further subdivide optimal control problems. These subdivisions are detailed in section 0.3.

0.2 The Statement of the Optimal Control Problem

Let the vector $x(u;t)$ denote the state of the system where the vector $u(t)$ is the control function and $J(u)$ is the cost functional.

The optimal control problem considered in this thesis can be stated as:

$$\text{maximize } J(u) = \int_0^1 L(x,u;t)dt$$

subject to

$$x'(u;t) = f(x,u;t)$$
$$x(u;o) = x_0 \qquad (0.1)$$

The spaces, norms, and restrictions for (0.1) will be presented in the ensuing chapters.

0.3 Classification of the Optimal Control Problem

This thesis considers four subdivisions of problem (0.1). They are - unconstrained, constrained, linear quadratic, and nonlinear.

0.3.1 Unconstrained - Unconstrained problems are maximizations of (0.1) over all control functions in its defining space.

0.3.2 Constrained - Constrained problems are maximizations of (0.1) over a subset of controls in its defining space.

0.3.3 Linear Quadratic - Linear quadratic problems have the following form,

$$\max J(u) = \tfrac{1}{2}\int_0^1 (x^T(t)Qx(t) + u^T(t)Ru(t))dt$$

subject to

$$x'(u;t) = Ax(u;t) + Bu(t)$$
$$x(u;o) = x_0 \qquad (0.2)$$

where Q is a symmetric negative semidefinite matrix and R is a symmetric negative definite matrix. In optimal control literature

$$x'(u:t) = A(t)x(u;t) + B(t)u(t)$$

is called a linear differential equation.

0.3.4 Nonlinear - Nonlinear problems are essentially described by (0.1); that is, the differential constraint is nonlinear.

0.4 The Maximum Principle of Pontryagin

Pontryagin et. al (15) proved that if there is an optimal control function v(t) which maximizes (0.1) then v(t) maximizes the Hamiltonian, $H(x,u,\lambda;t)$ where

$$H(x,u,\lambda;t) = f^T(x,u;t)\lambda^T(t),$$

f and x are defined in section 0.5.3 and

$$\lambda'(t) = -\frac{\partial H}{\partial x}(x,v,\lambda;t)\Big|_{x=x(v;t)}$$

$$\lambda_i(1) = 0 \qquad i=1,\ldots,n-1$$

$$\lambda_n(1)\geq 0$$

$\lambda(t)$ is called the adjoint or costate function. It can be considered as the dynamic equivalent of the Lagrangian for static constrained optimization problems. Thus, the maximum principle states that if v(t) maximizes (0.1), then

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

for all u(t) belonging to its defining space constrained or unconstrained where v(t) belongs to the same space or subset of the space.

0.5 Notation

0.5.1 Spaces -

$L_2^n\,[0,1]$	the Hilbert space of n component vector valued functions with inner product $(x\cdot y) = \int_0^1 x^T(t)y(t)dt$ such that $(x\cdot x)$ is integrable.
$PC_n[0,1]$	space of piecewise continuous n component vector valued functions defined on [0,1].

$C_n[0,1]$	the space of continuous n component vector valued functions defined on [0,1].
R^n	the Euclidean space of n component real valued vectors.

0.5.2 Norms -

$\|\|\cdot\|\|_2$	the $L_2^n[0,1]$ norm
$\|\|\cdot\|\|_M$	the maximum norm. If $x\varepsilon C_n[0,1]$
$\|\|x\|\|_M$	$\max_{1\le j\le n} \|\|x_j\|\|_M$
$\|\|\cdot\|\|_s$	the spectral norm. If $A\varepsilon$mxn real matrices
$\|\|A\|\|_s^2$	spectral radius of A^TA
$\|\|\cdot\|\|_n$	the n -dimensional Euclidean vector norm
$\|\|\cdot\|\|_\infty$	the essential supremum. If $x\varepsilon C_n[0,1]$ then, $\max_{1\le j\le n} \|\|x_j\|\|_\infty$

0.5.3 Functions - Let $f:R^n xR^m x[0,1]\rightarrow R^n$ where f is continous jointly in its first two arguments and piecewise continuous in its third argument.

$u(t)\varepsilon R^m, t\varepsilon[0,1]$denotes any control function while $v(t)\varepsilon R^m, t\varepsilon[0,1]$ denotes the optimal control.

$x_u(t) = x(u;t)\varepsilon R^n, t\varepsilon[0,1]$ denotes the function which describes the state of the system given any control function u.

$\lambda_u(t)\varepsilon R^n$ denotes the adjoint function given any control function u.

$$\frac{\partial H}{\partial H}_k(t) \equiv \frac{\partial H}{\partial u}(x_k, u_k, \lambda_k; t)$$

$$\frac{\partial H(u)}{\partial u} \equiv \frac{\partial H(x_u, u, \lambda_u; t)}{\partial u} .$$

0.5.4 Miscellaneous - A^T denotes the transpose of matrix A

$\overline{0}_n \equiv (0,\ldots,0)^T \varepsilon R^n$

$x^T Qx = (\sum_{i=1}^{n} x_i \sum_{j=1}^{n} q_{ij} x_j)$, $Q\varepsilon nxn$ matrices.

$N_v \equiv \{u : ||v-u||_2 < \varepsilon \text{ for some } \varepsilon > 0\}$.

Ω_A the set of admissible control functions

$Df(x,h) \equiv \lim_{t \to 0+} \frac{f(x + th) - f(t)}{t}$ if the limit exists.

$Df(x)h \equiv Df(x,h)$ when the right hand side is linear in h.

I_n the nxn identity matrix.

0_n the nxn zero matrix.

0.6 Integro - Differential Equations Associated with Optimal Control Problems

Much of what is presented here can be found in Bellman (1), Coddington and Levinson (3). Falb and deJong (6), or Halkin (7).

Let
$$x'(u;t) = f(x,u;t)$$
$$x(u;o) = x_o \qquad (0.3)$$
$$x(t)\varepsilon R^n \ , \ t\varepsilon[0,1]$$

where f is piecewise continuous with respect to $t\varepsilon[0,1]$, continuous with respect to x and $u, u\varepsilon L_2^m[0,1]$ and

$$||f(x,u;t)||_n \leq a(t)||u(t)||_m + b(t)||u(t)||_m||x(t)||_n$$

where a(t),b(t)>0 and integrable. For fixed essentially bounded u a solution to (0.3) is defined to be a function $\Phi(t)$ such that,

$$\Phi(t) = x_0 + \int_0^t f(\Phi,u;s)ds \quad \text{and}$$

$$\Phi'(t) = f(\Phi(t),u(t);t) \quad \text{a.e.}$$

The solution $\Phi(t)$ is absolutely continuous in $t\varepsilon[0,1]$.

Remark 0.1 (Coddington and Levinson (3)) There exists one and only one solution $\Phi(t)$ to (0.3) and it is bounded in [0,1] for each fixed u which is essentially bounded. In particular if

$$y'(t) = A(t)y(t) \quad \text{a.e.}$$

$$A(t)\varepsilon nxn \text{ matrices, } A_{ij}(t) \text{ measurable}$$

$$y(t)\varepsilon R^n \qquad (0.4)$$

$$t\varepsilon\ [0,1]$$

$||A(t)||_s \leq m(t)$ where m(t) is integrable, then there exists a unique solution Φ of (0.4) satisfying $\Phi(t_0) = \Phi_0$ and $\Phi'(t) = A(t)\Phi(t)$ a.e..

Remark 0.2 (Coddington and Levinson (3)) The set of all solutions to (0.4) form an n-dimensional vector space over R.

Let $y_1(t;t_0)$, $y_2(t;t_0),\ldots,y_n(t;t_0)$ be the set of n linearly independent solutions to (0.4) with initial condition at $t = t_0$. Define the fundamental matrix to be the matrix whose n columns are the n linearly independent solutions of (0.4). Let $\overline{\underline{Y}}^A(t;t_0)$ denote the fundamental matrix of (0.4) for which $\overline{\underline{Y}}^A(t_0,t_0) = I_n$. When the meaning is clear, $\overline{\underline{Y}}^A(t;t_0)$ will be written $\overline{\underline{Y}}(t;t_0)$.

Remark 0.3 For almost all $t\varepsilon[0,1]$

$$\frac{d}{dt}\underline{\bar{Y}}^A(t;t_0) = A(t)\underline{\bar{Y}}^A(t;t_0).$$

Let $\underline{\bar{Y}}(t;t_0) = \underline{\bar{Y}}^A(t;t_0)$ in the sequel

Remark 0.4 If

$$\frac{d}{dt}Z(t;t_0) = A(t)Z(t;t_0) \quad \text{a.e. and}$$

$$Z(t_0;t_0) = C$$

then, $Z(t;t_0) = \underline{\bar{Y}}(t;t_0)C$

Remark 0.5 (Coddington and Levinson (3)) $\underline{\bar{Y}}(t;t_0)$ is nonsingular in [0,1].

Remark 0.6 If

$$\frac{d}{dt}y(t) = A(t)y(t) \quad \text{a.e. and}$$

$$y(to) = C$$

then, $y(t) = \underline{\bar{Y}}(t;t_0)C.$

Remark 0.7 $\underline{\bar{Y}}^{-1}(t;t_0)$ satisfies the equation

$$\frac{d}{dt}Z(t;t_0) = -Z(t;t_0)A(t) \quad \text{a.e.}$$

$$Z(t_0,t_0) = I_n \qquad (0.5)$$

Proof

$$\frac{d}{dt}(I_n) = \frac{d}{dt}(\underline{\bar{Y}}(t;t_0)\underline{\bar{Y}}^{-1}(t;t_0))$$

$$0_n = \frac{d}{dt}\underline{\bar{Y}}(t;t_0)\underline{\bar{Y}}^{-1}(t;t_0) + \underline{\bar{Y}}(t;t_0)\frac{d}{dt}\underline{\bar{Y}}^{-1}(t;t_0)$$

$$-\underline{\bar{Y}}(t;t_0)\frac{d}{dt}\underline{\bar{Y}}^{-1}(t;t_0) = \frac{d}{dt}\underline{\bar{Y}}(t;t_0)\underline{\bar{Y}}^{-1}(t;t_0)$$

$$=A(t)$$

Therefore, $\frac{d}{dt}\underline{\bar{Y}}^{-1}(t;t_0) = -\underline{\bar{Y}}^{-1}(t;t_0)A(t)$ a.e.

which was to be shown.

Note that

$$\left[\frac{d}{dt}\underline{\overline{Y}}^{-1}(t;t_o)\right]^T = \frac{d}{dt}\underline{\overline{Y}}^{-1}{}^T(t;t_o) = -A^T(t)\underline{\overline{Y}}^{-1}{}^T(t;t_o)$$

$\underline{\overline{Y}}^{-1}(t;t_o)$ is called the adjoint matrix and it will be shown to be related to the adjoint function $\lambda(t)$ of the maximum principle.

Remark 0.8 $\underline{\overline{Y}}(s;t_1)\underline{\overline{Y}}(t_1;t) = \underline{\overline{Y}}(s;t)$

Proof Let

$$y'(t) = A(t)y(t) \quad \text{a.e.} \tag{0.6}$$

By remark 0.6, $y(t) = \underline{\overline{Y}}(t_1;t)y(t_1)$ unique. Thus,

$\underline{\overline{Y}}(s;t_1)\underline{\overline{Y}}(t_1;t)y(t) = y(s)$ for all $s,t_1,t\varepsilon[0,1]$.

But $\underline{\overline{Y}}(s;t)y(t) = y(s)$. Therefore,

$\underline{\overline{Y}}(t;t_1)\underline{\overline{Y}}(t_1;t)y(t) = \underline{\overline{Y}}(s;t)y(t)$ for all y satisfying 0.4.

Thus, $\underline{\overline{Y}}(t;t_1)\underline{\overline{Y}}(t_1;t) = \underline{\overline{Y}}(s;t)$

Remark 0.9 $\underline{\overline{Y}}(t;s) = \underline{\overline{Y}}^{-1}(s;t)$

Proof: From remark 0.8, let $t = s$

$$\underline{\overline{Y}}(s;t_1)\underline{\overline{Y}}(t_1;s) = \underline{\overline{Y}}(s,s) = I_n$$

which implies

$$\underline{\overline{Y}}(t_1;s) = \underline{\overline{Y}}^{-1}(s;t_1)$$

But t_1 is arbitrary, and the remark is proved.

Remark 0.10 Let $y'(t) = A(t)y(t) + p(t)$ a.e.

$$y(0) = c$$

Then, $y(t) = \underline{\overline{Y}}(t;0)c + \int_0^t \underline{\overline{Y}}^{-1}(s;t)p(s)ds$

Proof: $y(t) = \underline{\overline{Y}}(t;0)c + \int_0^t \underline{\overline{Y}}(t;0)\underline{\overline{Y}}^{-1}(s;0)p(s)ds$

$$= \underline{\overline{Y}}(t;0)c + \int_0^t \underline{\overline{Y}}(t;0)\underline{\overline{Y}}(0;s)p(s)ds$$

$$= \underline{\overline{Y}}(t;0)c + \int_0^t \underline{\overline{Y}}(t;s)p(s)ds$$

$$= \underline{\overline{Y}}(t;0)c + \int_0^t \underline{\overline{Y}}^{-1}(s;t)p(s)ds$$

Remark 0.11 (Bellman (1)). If

$$\frac{d}{dt}\underline{\overline{Y}}(t;o) = A\underline{\overline{Y}}(t;o) \quad \text{a.e.}$$

$$\underline{\overline{Y}}(o;o) = I_n$$

then $\underline{\overline{Y}}(t;o) = e^{At}$ where

$$e^{At} \equiv I_n + At + \frac{A^2t^2}{2} + \ldots + \frac{A^nt^n}{n!} + \ldots \qquad (0.9)$$

Remark 0.12 (Bellman (1)) The matrix series (0.9) for e^{At} exists for all A for any fixed value of t and for all values of t for any fixed A. It converges uniformly in any finite region of the complex t plane.

Remark 0.13 If

$$\frac{d}{dt} y(t) = Ay(t) + p(t)$$

$$y(o) = c$$

then $y(t) = e^{At}c + \int_0^t e^{A(t-s)}p(s)ds$

Remark 0.14 (a) $[\int A(t)dt]^T = \int A^T(t)dt$

(b) $[e^{At}]^T = e^{A^Tt}$

0.7 Interval Analysis - References and Definitions

Guaranteed error bounds are computed using interval analysis. A basic textbook on the subject of interval analysis is R.E. Moore (13). L.T. Winter (17) has an introduction to interval analysis. The interval analysis notation of Winter (17) is used in this thesis. In particular,

$$a^I \equiv \{x : a^- \leq x \leq a^+\} \text{ or } a^I = [a^-, a^+]$$

An interval valued function f^I will be called an interval extension of f on a^I if, for every $b^I \subset a^I$, $f^I(b^I)$ is defined and $\{f(t) : t \varepsilon b^I\} \subset f^I(b^I)$. This concept extends to real functions of more than one variable by allowing each argument to range in an interval. Intervals are connected, compact subsets of R.

Definition 0.1 An optimal control problem (0.1) is called normal if

$$\frac{\partial H}{\partial u}(x,u,\lambda;t) = o$$

can be solved for u in terms of x, λ, and t. It is locally normal if this is true in a neighborhood of v, x_v, and λ_v.

1. THE PROOF OF THE MAXIMUM PRINCIPLE OF PONTRYAGIN

1.1 Proof

The maximum principle of Pontryagin or simply the maximum principle is proved here using Fréchet derivatives. Thus, an extension of the usual calculus methods for finding relative maxima is obtained for optimal control problems. The Fréchet derivative approach yields necessary conditions for local maxima; however, these necessary conditions are identical to those obtained by Pontryagin et. al. (15).

The proof of the maximum principle contained here is a weaker result than found in Pontryagin (15) only because the differential constraint must be assumed to be concave. Otherwise, the results contained here are the same as found in Pontryagin (15). The proof is simple and well motivated in that it generalizes the ordinary calculus notion of taking a derivative and setting it equal to zero to find a relative maximum or minimum.

The results (intermediate and final) of the maximum principle as proved here are used in the sequel. That is the equations, intermediate and final, are used in chapters 2, 3 and 4. These are used in developing the two numerical methods derived here to obtain approximate solutions with guaranteed error bounds.

Explicit solutions using the maximum principle have been obtained for a small class of problems. In particular, linear quadratic problems have explicit solutions in the unconstrained case. Most optimal control problems require numerical methods for obtaining (approximate) solutions.

Theorem 1.1 (The Maximum Principle) Let $\Omega_A^1 \equiv \{u : u \varepsilon L_2^m[0,1], ||u||_\infty \leq 1\}, \Omega_A^2 = \{u : u \varepsilon L_2^m[0,1], ||u||_\infty \leq K_u$, K_u a finite number for each u} and Ω_A is either Ω_A^1 or Ω_A^2.

$$J(u) = \int_o^1 L(x,u;t)dt$$

$$x(u):[0,1] \to R^n$$

$$x'(u;t) = f(x,u;t) \qquad \text{such that}$$

$$x_n'(u;t) = f_n(x,u;t) \equiv L(x,u;t) \qquad (1.1)$$

$$x(u;o) = x_0 \qquad \text{where}$$

$$x_n(u;o) = 0$$

$x_1,\ldots,x_{n-1}$ are the states of the system and $x_n(u;1) = J(u)$ is the cost functional.

Let $D_i = R^n x R^m x (t_{i-1}, t_i)$ $i = 1, \ldots, N<\infty$ $0=t_0<t_1<\ldots<t_N = 1$

and $O(f) = (t_0, t_1) \cup \ldots \cup (t_{N-1}, t_N)$. Define

$$f^i(x,u;t) = f(x,u;t) \quad \text{if } t\varepsilon(t_{i-1},t_i) \text{ where}$$

(i) $f^i: D_i \to R^n$ is continuous

(ii) $\frac{\partial f^i}{\partial x}: D_i \to R^n x R^n$ is continuous (nxn matrix)

(iii) $\frac{\partial f^i}{\partial u}: D_i \to R^n x R^m$ is continuous (nxm matrix)

and (iv) $||f^i(x,u;t)||_n \leq a(t)||u(t)||_m + b(t)||u(t)||_m ||x(u;t)||_n$ for $a(t), b(t)>0$ and integrable. Given $(x(t),u(t)) \subset B$ any bounded set in $R^n x R^m$ let there exist a positive constant L_B such that

$$||\frac{\partial f^i}{\partial x}(x_1,u_1;t)-\frac{\partial f^i}{\partial x}(x_2,u_2;t)||_s \leq L_B(||x_1-x_2||_n^2 + ||u_1-u_2||_m^2)^{\frac{1}{2}}$$

and

$$||\frac{\partial f^i}{\partial u}(x_1,u_1;t)-\frac{\partial f^i}{\partial u}(x_2,u_2;t)||_s \leq L_B(||x_1-x_2||_n^2 + ||u_1-u_2||_n^2)^{\frac{1}{2}}$$

where L_B is independent of x, u and t but not necessarily of the set B. Lastly, let $f(x,u;t), t\varepsilon O(f)$ be concave as a function of u in N_v, $v\varepsilon\Omega_A$ where v is assumed to exist and

$$x_n(v;1) \geq x_n(u;1) \qquad \text{for all } u\varepsilon\Omega_A \cap N_v$$

Then there exists a vector valued function $\lambda(t), \lambda:[0,1] \to R^n$ defined and continuous over [0,1], differentiable on O(f) and not identically zero such that,

(1) $H(x_v,v,\lambda_v;t) \geq H(x_v,\mu,\lambda_v;t)$ for all $u\varepsilon\Omega_A \cap N_v$

(2) $\lambda'(t) = - \frac{\partial H}{\partial x}(x_v,v,\lambda_v;t)$

(3) $\lambda_i(1) = 0 \qquad i=1,\ldots,n-1$

(4) $\lambda_n(1) = 1$

Before proving the maximum principle a few lemmata and remarks are presented. This proof of the maximum principle simplifies the proof by Halkin (7) though Halkin, like Pontryagin, obtains global results. In addition, Halkin (7) considers right hand endpoint constraints for as many as the first n-2 states. The set Ω_A in Halkin (7) is the set of $u\varepsilon\ PC_m[0,1]$ such that $u:[0,1]\to\Omega\subset R^m$ where Ω is essentially bounded. Thus, Ω_A in this thesis is more general.

The proof of the maximum principle given in Hasdorff (9) is typical of many of the intuitive approaches to the maximum principle. Though he uses Fréchet derivatives in the proof, only the unconstrained case is handled. The proof of the maximum principle contained in this thesis is rigorous and considers the constrained case when Ω_A is Ω_A^1 and the unconstrained case when Ω_A is Ω_A^2.

Remark 1.2 Let Ω be bounded in $L_2^m[0,1]$, $u\varepsilon\Omega$. Suppose $f(x,u;t)$ satisfies the conditions for f in theorem 1.1. Then the set of vectors $\{x(u;t):u\varepsilon\Omega, t\varepsilon[0,1]\}$ is bounded in R^n.

Proof $||f(x,u;t)||_n \leq a(t)||u(t)||_m + b(t)||u(t)||_m||x(u;t)||_n$

Define $N(x) = ||x||_n$, $\Phi(t) = ||x(t)||_n$. Then using the fact that norms are differentiable since they are convex, the chain rule, and the fact that $DN(x,h)\leq N(h)$

$$\Phi'(t) = DN(x(t),x'(t))\leq a(t)||u(t)||_m + b(t)||u(t)||_m\ \Phi(t)\ .$$

Let $\quad B(t) = \int_0^t b(s)||u(s)||_m ds$, then

$$(\Phi'(t) - a(t)||u(t)||_n - B'(t)\ \Phi(t))e^{-B(t)}\leq 0 \quad \text{a.e.}\ .$$

That is $\quad \Phi(t)e^{-B(t)} - \int_0^t a(s)||u(s)||_m e^{-B(s)}ds$ is nonincreasing

Thus, $\quad \Phi(t)e^{-B(t)} - \int_0^t a(s)||u(s)||_m e^{-B(s)}ds\leq\Phi(o) = ||x_o||_n$.

And $\Phi(t) \leq ||x_0||_n e^{B(t)} + e^{B(t)} \int_0^t a(s)||u(s)||_m e^{-B(s)} ds$.

But $\int_0^t b(s)||u(s)||_m ds \leq \int_0^1 b(s)||u(s)||_m ds$

$$\leq ||b||_2 \cdot ||u||_2$$

$$\leq B_1 \cdot K \ .$$

Thus, $\int_0^t a(s)||u(s)||_m e^{-B(s)} ds \leq A_1 K \cdot e^{B_1 K}$

Thus Φ is bounded $t \varepsilon [0,1]$ which was to be shown.

If Ω_A is Ω_A^1, then $x(u;t)$ is bounded. If Ω_A is Ω_A^2, then $x(u;t)$ is bounded for each u. Thus, the set B of theorem 1.1 can be considered as being generated by the essentially bounded functions u of Ω_A^1 or Ω_A^2. In the latter case, L_B is a constant depending only on u.

Let $\Omega = \{u: ||u||_\infty \leq K_u, K_u$ finite, $u \varepsilon L_2^m[0,1]\}$. Then for $u \varepsilon \Omega$, $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial u}$ are bounded since they satisfy a Lipschitz condition on bounded sets of $(x(t),u(t))$.

Remark 1.3 Ω_A^1 is a convex, weakly compact subset of $L_2^m[0,1]$.

Proof That Ω_A^1 is convex follows immediately from its definition. Since Ω_A^1 is convex, it is sufficient to show Ω_A^1 is closed and bounded.

Let $u_n \ \varepsilon \Omega_A^1$ and $u_n \to u$ in $L_2^m[0,1]$, then $u \varepsilon L_2^m[0,1]$. Rudin (15) page 70 proves that there exists a subsequence of $\{u_n\}$ which converges pointwise almost everywhere to $u(t)$. Let u_{n_j} be that subsequence. Since norms are continuous

$$||u_{n_j}(t)||_m \to ||u(t)||_m \quad \text{a.e.}$$

But $||u_{n_j}(t)||_m \leq 1$ a.e., therefore $||u(t)||_m \leq 1$ a.e. so that $||u||_\infty \leq 1$ and $u \varepsilon \Omega_A^1$. Therefore, Ω_A^1 is closed. It is obvious that Ω_A^1 is bounded $L_m^2[0,1]$. Thus, the remark is proved.

Remark 1.4 Let $u_1, u_2 \varepsilon \Omega_A$. Let C be any measurable subset of $[0,1]$. Define u_3 as follows:

$$u_3(t) = \begin{cases} u_1(t) & t\varepsilon C \\ u_2(t) & t\varepsilon[0,1]\setminus C \end{cases}$$

Then $u_3 \varepsilon \Omega_A$.

<u>Proof</u> $||u_3(t)||_m \leq \max \{||u_1(t)||_m , ||u_2(t)||_m\}$

Obviously, $u_3 \varepsilon L_2^m[0,1]$. Therefore, $u_3 \; \varepsilon\Omega_A$ which was to be shown.

The set $\{u:||u||_2 \leq 1\}$ does not satisfy the above remark.

To see this, let

$$u_1(t) = \begin{cases} 2 & 0\leq t\leq \frac{1}{2} \\ 0 & \frac{1}{2}<t\leq 1 \end{cases}$$

$$u_2(t) = \begin{cases} 0 & 0\leq t<\frac{1}{2} \\ 2 & \frac{1}{2}\leq t\leq 1 \end{cases}$$

As will be seen in the proof of the maximum principle, it is essential that controls have the property described in remark 1.4.

<u>Remark 1.5</u> The following theorems are stated without proofs since the proofs can be found in standard text books on optimization or approximation theory (see Demynov and Rubinov (4)).

Let f be a functional from a set $S\subset \underline{X}$ to the extended reals R_∞.

(1) f concave on S implies that there exists $Df(x,h)$, $h\varepsilon S-x$. If $Df(x^*,h) \leq 0$ for all $h\varepsilon S-x^*$, then

$$f(x^*) = \sup_{x\varepsilon S} f(x) \; .$$

(2) Let S be a convex set. If $f(x^*) = \sup_{x\varepsilon S} f(x)$ and if $Df(x^*,h)$ exists, $h\varepsilon S-x^*$ then

$$Df(x^*,h)\leq 0 \qquad \text{for all } h\varepsilon S-x^* \; .$$

(3) If $f(x^*) = \sup_{x\varepsilon S} f(x)$ where x^* belongs to the interior of the

set S and if there exists $Df(x)h$, $h \in S-x$ for all x in a neighborhood of x^*, then

$$Df(x^*)h = 0.$$

Gronwall's Inequality Let K be a constant, f(t) a continuous function and g(t) an integrable nonnegative function for $a \leq t \leq b$ satisfying the inequality

$$f(t) \leq K + \int_a^t g(s)f(s)ds ,$$

then $$f(t) \leq K \cdot \exp(\int_a^t g(s)ds) \qquad a \leq t \leq b.$$

Proof Let $p(t) = \int_a^t g(s)ds$ and

$$r(t) = \int_a^t g(s)f(s)ds$$

$$g(t)f(t) \leq g(t)K + g(t)(\int_a^t g(s)f(s)ds)$$

$$r'(t) - g(t)K - g(t)r(t) \leq 0$$

$$\frac{d}{dt}(e^{-p(t)}(r(t) + K)) \leq 0$$ so that

$e^{-p(t)}(r(t) + K)$ is a nonincreasing function on the interval $[a,b]$. Thus

$$e^{-p(t)}(r(t) + K) \leq e^{-p(a)}(r(a) + K) = K$$

$$r(t) + K \leq K \cdot e^{p(t)} .$$

But $$r(t) + K = \int_a^t g(s)f(s)ds + K \geq f(t)$$ implies that

$$f(t) \leq K \exp (\int_a^t g(s)ds) \qquad a \leq t \leq b$$

which was to be shown.

Remark 1.6 Let $f: S \subset \underline{X} \to Z$ be given where S is convex, $\underline{X}$ and Z are Banach spaces. If f is differentiable on $E = [0,1] \setminus \{t_1, \ldots, t_n\}$, then

$$||f(x_1) - f(x_2)||_Z \leq \sup_{t \in E} ||Df(tx_1 + (1-t)x_2, (x_1 - x_2))||_Z .$$

The proof to the above remark can be found in Dieudonné (5).

Remark 1.7 Lef $f\mathcal{S}\subset\underline{\overline{X}}$ Z be given where S is convex and $\underline{\overline{X}}$ and Z are Banach spaces. Let f(x) be Fréchet differentiable such that Df(x)h has a Lipschitz constant M_S with respect to x. Then,

$$||f(x + w) - (f(x) + Df(x)w)||_Z \leq M_S ||w||^2_{\underline{\overline{X}}} \quad . \quad (1.2)$$

Proof Let $\Phi(t) = f(x + w) - f(x) - tDf(x)w$. Using remark 1.6

$$||f(x + w) - (f(x) + Df(x)w)||_Z = ||\Phi(1)||_Z$$

$$\leq \sup_{0\leq t\leq 1} ||D\Phi(t)||_Z$$

$$= \sup_{0\leq t\leq 1} ||Df(x + tw)w - Df(x)w||_Z$$

$$\leq \sup_{0\leq t\leq 1} ||Df(x + tw) - Df(x)||_Z ||w||_{\underline{\overline{X}}}$$

$$\leq \sup_{0\leq t\leq 1} tM_S ||w||^2_{\underline{\overline{X}}}$$

$$= M_S ||w||^2_{\underline{\overline{X}}}$$

which was to be shown.

By partitioning [0,1] into small subintervals and taking limits, the above inequality can be "improved" by a factor of ½ and it can be shown that Ms/2 is the smallest factor that can be obtained. For this thesis (1.2) will suffice.

Remark 1.8 Let $f:W = \underline{\overline{X}} x \underline{\overline{Y}} \to Z$ be given where $\underline{\overline{X}}$,$\underline{\overline{Y}}$ and Z are Banach spaces and $||\cdot||_W \equiv (||\cdot||^2_{\underline{\overline{X}}} + ||\cdot||^2_{\underline{\overline{Y}}})^{\frac{1}{2}}$.

If $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist and are continuous, then for any $w_0 = (x_0,y_0)\varepsilon W$, $\ell = (h,k)\varepsilon W$

$$Df(w_0)\ell = \frac{\partial f}{\partial x}(x_0,y_0)h + \frac{\partial f}{\partial y}(x_0,y_0)k.$$

The proof of the above remark can be found in Dieudonné (5).

Remark 1.9 Let $f: S \subset W = \overline{X} \times \overline{Y} \to Z$ be given where $\overline{X}, \overline{Y}, Z$ and $||\cdot||_W$ are defined as above, S a convex bounded set. Let $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ have Lipschitz constants of M_{s1}, M_{s2} respectively on S with respect to x and y. Then the derivative of f has a Lipschitz constant of $M_s = M_{s1} + M_{s2}$.

Proof Let $w = (x,y) \varepsilon S$ and $\ell = (h,k) \varepsilon S$.

$$||Df(w_1) - Df(w_2)|| = \max_{||\ell||_W \leq 1} ||Df(w_1)\ell - Df(w_2)\ell||_Z$$

$$\leq \max_{||\ell||_W \leq 1} ||\frac{\partial f}{\partial x}(x_1,y_1)h + \frac{\partial f}{\partial y}(x_1,y_1)k -$$

$$\frac{\partial f}{\partial x}(x_2,y_2)h - \frac{\partial f}{\partial y}(x_2,y_2)k||_Z$$

$$\leq \max_{||\ell||_W \leq 1} (M_{s1}(||x_1-x_2||^2_{\overline{X}} + ||y_1-y_2||^2_{\overline{Y}})^{\frac{1}{2}}||h||_{\overline{X}} +$$

$$M_{s2}(||x_1 - x_2||^2_{\overline{X}} + ||y_1 - y_2||^2_{\overline{Y}})^{\frac{1}{2}}||k||_{\overline{Y}})$$

$$\leq (M_{s1} + M_{s2})||w_1 - w_2||_W$$

which was to be shown.

Remark 1.10 Let $f: S \subset W = \overline{X} \times \overline{Y} \to Z$ be defined and satisfy the conditions of remark 1.9. Then

$$||f(x+h,y+k) - (f(x,y) + \frac{\partial f}{\partial x}(x,y)h + \frac{\partial f}{\partial y}(x,y)k)||_Z \leq M_s(||h||^2_{\overline{X}} + ||k||^2_{\overline{Y}}) . \qquad (1.3)$$

Proof From equation (1.2)

$$||f(w) - (f(u) + Df(u)\ell)||_Z \leq M_s||\ell||^2_W \text{ where}$$

$w = (x+k, y+k), u = (x,y)$ and $\ell = (h,k)$. $M_s = M_{s1} + M_{s2}$ by remark 1.9. From remark 1.8,

$$||f(w) - (f(u) + Df(u)\ \ell)||_z =$$

$$||f(x+h,y+k) - (f(x,y) + \frac{\partial f}{\partial x}(x,y)h + \frac{\partial f}{\partial y}(x,y)k)|| \leq M_s ||\ell||_w^2.$$

Since $||\ell||_w^2 = ||h||_{\underline{x}}^2 + ||k||_{\underline{y}}^2$, the remark is proved.

Let G(u) denote the solution to the differential equation constraint of theorem 1.1,

$$x'(u;t) = f(x,u;t)$$

$$x(u;o) = x_o \quad .$$

By the assumptions to theorem 1.1, G(u) exists, is unique, continuous on O(f) and bounded on [0,1] for each $u \varepsilon \Omega_A$.

$G:\Omega_A \to C_n[0,1]$ where the norm of $C_n[0,1]$ will be the maximum norm $||\cdot||_M$.

Lemma 1.11 Let the conditions for theorem 1.1 hold. Then G(u) is uniformly continuous on Ω_A^1 and on Ω any bounded convex subset of Ω_A^2.

Proof Let $x(u;t) = G(u)$ for a fixed $u\varepsilon\Omega$. Here Ω will be as above or Ω_A^1. It must be shown that $y(w;t) \to x(u;t)$ in the maximum norm as $h \equiv w-u \to o$ in the $L_2^m[0,1]$ norm. For all $t\varepsilon[0,1]$, setting $z(t) = y(w;t) - x(u;t)$ the following holds.

$$||z(t)||_n = ||y(w;t) - x(u;t)||_n$$

$$= ||\int_o^t (f(y,w;s) - f(x,u;s))ds||_n$$

$$\leq \int_o^t ||(f(y,w;s) - f(x,w;s) + f(x,w;s) - f(x,u;s)||_n ds$$

$$\leq \int_o^t ||f(y,w;s) - f(x,w;s)||_n ds +$$

$$+ \int_o^t ||f(x,w;s)-f(x,u;s)||_n ds$$

by remark 1.6 $\leq \int_0^t \sup_{0\leq\alpha\leq 1} \left|\left|\frac{\partial f}{\partial x}(\alpha y+(1-\alpha)x,w;s)z(s)\right|\right|_n ds +$

$\int_0^t \sup_{0\leq\beta\leq 1}\left|\left|\frac{\partial f}{\partial u}(x,\beta w+(1-\beta)u;s)h(s)\right|\right|_n ds$

$\leq \int_0^t \sup_{0\leq\alpha\leq 1} \left|\left|\frac{\partial f}{\partial x}(\alpha y+(1-\alpha)x,w;s\right|\right|_s ||z(s)||_n ds +$

$\int_0^t \sup_{0\leq\beta\leq 1} \left|\left|\frac{\partial f}{\partial u}(x,\beta w+(1-\beta)u;s\right|\right|_s ||h(s)||_m ds$.

By remark 1.2 $\left|\left|\frac{\partial f}{\partial x}\right|\right|_s \leq K_1$ and $\left|\left|\frac{\partial f}{\partial u}\right|\right|_s \leq K_2$ where K_1 and K_2 depend on Ω. Therefore

$$||z(t)||_n \leq K_2 \int_0^1 ||h(s)||_m ds + \int_0^t K_1 ||z(s)||_n ds$$

Applying Hölder's inequality to the first integral,

$$||z(t)||_n \leq K_2 ||h||_2 + \int_0^t K_1 ||z(s)||_n ds$$

Applying Gronwall's inequality to the above,

$$||z(t)||_n \leq K_2 ||h||_2 e^{K_1 t}$$

$\leq K||h||_2$. Since the right hand side does not depend on t,

$||z||_M \leq K||h||_2$ which was to be shown. Note that lemma 1.11 implies that $\frac{\partial f}{\partial x}$, and $\frac{\partial f}{\partial u}$ satisfy a Lipschitz condition with respect to $u\varepsilon\Omega_A$.

<u>Lemma 1.12</u> Let the conditions to theorem 1.1 hold. Then G(u) is Fréchet differentiable for all $u\varepsilon\Omega_A^1$ or for all $u\varepsilon\Omega=\Omega_A^2$, Ω <u>convex</u> bounded, and

$$DG(u)h = \int_0^t \overline{Y}^{-1}(s;t)\frac{\partial f}{\partial u}(x,u;s)h(s)ds \qquad (1.4)$$

where $\overline{Y}(t;o)$ is the fundamental matrix of

$$q'(u;t) = \frac{\partial f}{\partial x}(x,u;t)q(u;t)$$

$h\varepsilon\Omega_A - u$, Ω_A is Ω_A^1 or Ω.

<u>Proof</u> Let h and z be defined as in lemma 1.11.

$$z'(t) = f(y,w;t) - f(x,u;t)$$

$$= \frac{\partial f}{\partial x}(x,u;t)z(t) + \frac{\partial f}{\partial u}(x,u;f)h(t) + R(x,u,z,h;t) .$$

Let $R(t) \equiv R(x,u,z,h;t)$. By remark 1.10,

$$||R(t)||_n = ||f(y,w;t) - f(x,u;t) - \frac{\partial f}{\partial x}(x,u;t)z(t) - \frac{\partial f}{\partial u}(x,u;t)h(t)||_n$$

$$\leq 2L(||z(t)||_n^2 + ||h||_2^2)$$

$$\leq 2L(||z||_M^2 + ||h||_2^2)$$

where L is the Lipschitz constant for $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial u}$ on Ω_A. L depends on Ω_A. By lemma 1.11,

$$||R(t)||_n \leq 2L(K^2||h||_2^2 + ||h||_2^2)$$

$$= 2L(K^2+1)||h||_2^2 = M||h||_2^2 .$$

Let $p'(u;t) = \frac{\partial f}{\partial x}(x,u;t)p(u;t) + \frac{\partial f}{\partial u}(x,u;t)h(t)$

$p(u;o) = 0$

Let $E(t) = ||z(t) - p(u;t)||_n$

$$= ||\int_0^t \frac{\partial f}{\partial x}(x,u;s)(z(s) - p(u;s))ds + \int_0^t R(s)ds||_n$$

$$\int_0^t ||\frac{\partial f}{\partial x}(x,u;s)||_s E(s)ds + M||h||_2^2 .$$

Using Gronwall's inequality and recalling that $\frac{\partial f}{\partial x}$ is bounded by K_1 on Ω_A,

$$E(t) \leq M \cdot ||h||_2^2 e^{M_1 t}$$

$$\leq \overline{M}||h||_2^2$$

But $G(u+h) - G(u) = z(t)$. If $p(u;t)$ is linear in h and bounded on Ω_A, then it is the Fréchet derivative of G at u in the direction h. From remarks 0.8, 0.9, and 0.10,

$$p(u;t) = \int_0^t \underline{\overline{Y}}(t;o)\underline{\overline{Y}}^{-1}(s;o)\frac{\partial f}{\partial u}(x,u;s)h(s)ds$$

$$= \int_0^t \underline{\overline{Y}}^{-1}(s;t)\frac{\partial f}{\partial u}(x,u;s)h(s)ds\,.$$

Thus p(u;t) is linear in h. Writing p(u;t) as an integral equation,

$$p(u;t) = \int_0^t \frac{\partial f}{\partial x}(x,u;s)p(u,s)ds + \int_0^t \frac{\partial f}{\partial u}(x,u;s)h(s)ds$$

Proceeding as before and using Gronwall's inequality, it can be seen that

$$||p||_M \leq ||\frac{\partial f}{\partial u}||_s ||h||_2 \exp(||\frac{\partial f}{\partial x}||_s)$$

$$\leq K||h||_2 \quad \text{where } K = K_2 e^{K_1}\,.$$

Since Ω_A is bounded so is Ω_A-u and p is bounded, which was to be shown.

Let $\underline{\overline{Y}}(t;o)$ denote the fundamental matrix of lemma 1.12 and Ω_A be Ω_A^1 or $\Omega c\ \Omega_A^2$.

Remark 1.13 $\underline{\overline{Y}}(t;o)$ satisfies a Lipschitz condition in u for $u\varepsilon\Omega_A$.

$$\frac{d}{dt}\underline{\overline{Y}}(t;o) = \frac{\partial f}{\partial x}(x,u;t)\underline{\overline{Y}}(t;o) \qquad \underline{\overline{Y}}(o;o) = I_n$$

Let $F(x,u,\underline{\overline{Y}};t) = \frac{\partial f}{\partial x}(x,u;t)\underline{\overline{Y}}(t;o)$. Then

$\underline{\overline{Y}}(t;o) = I_n + \int_0^t F(x,u,\underline{\overline{Y}};s)ds$. Let $y(t) = (x(t),u(t))$ and $W = R^n x R^m$, $||\cdot||_W$ as before. $F(x,u,\underline{\overline{Y}};t)$ satisfies a Lipschitz condition in y and $\underline{\overline{Y}}(t;o)$ since

$$||F(y_1,\underline{\overline{Y}};t) - F(y_2,\underline{\overline{Y}};t)||_s \leq ||\frac{\partial f}{\partial x}(y_1;t) - \frac{\partial f}{\partial x}(y_2;t)||_s ||\underline{\overline{Y}}(t;o)||_s$$

$$\leq L_1||y_1-y_2||_W||\underline{\overline{Y}}(t;o)||_s\,.$$

The last inequality is true since $\frac{\partial f}{\partial x}$ is Lipschitz on Ω_A where L_1 is its Lipschitz constant.

$$||F(y,\underline{\overline{Y}}_1;t) - F(y,\underline{\overline{Y}}_2;t)||_s \leq ||\frac{\partial f}{\partial x}(y;t)||_s||\underline{\overline{Y}}_1(t;o) - \underline{\overline{Y}}_2(t;o)||_s$$

where $\underline{\overline{Y}}_1(t;o)$ is the fundamental matrix for

$q^1(u_1;t) = \frac{\partial f}{\partial x}(x_1,u_1;t)q(u_1;t)$ and

$\underline{Y}_2(t;o)$ for $q^1(u_2;t) = \frac{\partial f}{\partial x}(x_2,u_2;t)q(u_2;t)$. But $||\frac{\partial f}{\partial x}||_s \leq K_1$ on Ω_A and $F(y,\underline{Y};t)$ is Lipschitz in $\underline{Y}$. Let L_2 be its Lipschitz constant. By remark 1.9,

$$||F(y_1,\underline{Y}_1;t) - F(y_2,\underline{Y}_2;t)||_2^2 \leq (L_1+L_2)(||y_1-y_2||_w^2+||\underline{Y}_1-\underline{Y}_2||_2^2).$$

Thus, letting $L = L_1 + L_2$

$$||\underline{Y}_1(t;o) - \underline{Y}_2(t;o)||_s^2 \leq \int_0^t ||F(y_1,\underline{Y}_1;s) - F(y_2,\underline{Y}_2;s)||_s^2 ds$$

$$\leq L\int_0^t ||y_1(s) - y_2(s)||_w^2 ds + \int_0^t L||\underline{Y}_1(s;o) - \underline{Y}_2(s;o)||_s^2 ds.$$

But $||y_1(s) - y_2(s)||_w^2 = ||x_1(s) - x_2(s)||_n^2 + ||u_1(s) - u_2(s)||_m^2$

by lemma 1.11 $\leq K||u_1 - u_2||_2^2 + ||u_1(s) - u_2(s)||_m^2$

$\leq (K+1)||u_1 - u_2||_2^2.$

Thus $||\underline{Y}_1(t;o) - \underline{Y}_2(t;o)||_s^2 \leq L(K+1)||u_1 - u_2||_2^2 +$

$\int_0^t L||\underline{Y}_1(s;o) - \underline{Y}_2(s;o)||_s ds$.

Using Gronwall's inequality and letting M^2 be the resulting constant,

$$||\underline{Y}_1(t;o) - \underline{Y}(t;o)||_s^2 \leq M^2||u_1 - u_2||_2^2$$ which was to be shown.

Note that $\underline{Y}^{-1}(s;o)$ is Lipschitz in $u\varepsilon\Omega_A$ since by remark 0.7

$$[\frac{d}{dt}\underline{Y}^{-1}(s;o)]^T = \frac{d}{ds}\underline{Y}^{T-1}(s;o) = -\frac{\partial f}{\partial x}^T(x,u;s)\underline{Y}^{T-1}(s;o)$$

<u>Remark 1.14</u> If $f(x)$ and $g(x)$ satisfy a Lipschitz condition on a bounded set $S\subset\underline{X}$, then $f(x)g(x)$ and $f(g(x))$ also satisfy a Lipschitz condition on $S\subset\underline{X}$ with appropriate domains and ranges.

Note that remark 1.13 and 1.14 imply that $\underline{Y}^{-1}(s;t) = \underline{Y}(t;o)\underline{Y}^{-1}(s;o)$ satisfies a Lipschitz in $u\varepsilon\Omega_A$.

Let $\Phi:C_n[0,1]\to R$ be defined by

$$\Phi(x) \equiv x_n(u;1) \qquad (1.5)$$

<u>Remark 1.15</u> Φ is linear and the operator norm is 1.

Remark 1.16 $\Phi(G(u)) = J(u)$

Remark 1.17 $DJ(u)h = \Phi(DG(u)h)$

Remark 1.18 Let the conditions to theorem 1.1 hold. Then $DG(u)h$ satisfies a Lipschitz condition on Ω_A.

Proof $DG(u)h = \int_0^t \underline{\overline{Y}}^{-1}(s;t)\, \frac{\partial f(x,u;s)}{\partial u} h(s)ds$ by (1.4). $\underline{\overline{Y}}^{-1}(s;t)$ and $\frac{\partial f(x,u;s)}{\partial u}$ satisfy Lipschitz conditions with respect to u, the latter by the hypothesis to theorem 1.1 and lemma 1.11 since u is essentially bounded. Therefore, by remark 1.14 the product is Lipschitz and the remark is proved.

Remark 1.19 Let the conditions for theorem 1.1 be satisfied. The Φ of equation (1.5) satisfies a Lipschitz condition on Ω_A.

Proof
$$|\Phi(x_1) - \Phi(x_2)| = |x_n(u_1;1) - x_n(u_2;1)|$$
$$\leq K||u_1 - u_2||_2$$
by lemma 1.11 and the remark is proved.

Remark 1.20 Let the conditions for theorem 1.1 hold. Then $DJ(u)h$ satisfies a Lipschitz condition with respect to u in Ω_A.

Proof Since $DJ(u)h = \Phi(DG(u)h)$, remarks 1.18, 1.19, and 1.14 imply that $DJ(u)h$ is Lipschitz which was to be shown.

Proof (of the maximum principle)

By remark 1.17, lemma 1.12, and (1.5)

$$DJ(v)h = \Phi(DG(v)h)$$
$$= \Phi\left(\int_0^1 \underline{\overline{Y}}^{-1}(s;t)\, \frac{\partial f(x_v,v;s)}{\partial u} h(s)ds\right)$$
$$= \int_0^1 \underline{\overline{Y}}_n^{-1}(s;1) \frac{\partial f(x_v,v;s)}{\partial u} h(s)ds \qquad (1.6)$$

where $\underline{\overline{Y}}_n^{-1}(s;t)$ denotes the n^{th} row of $\underline{\overline{Y}}^{-1}(s;t)$. By remark 0.7

$$\frac{d}{ds}\underline{\overline{Y}}^{-1}(s;1) = -\underline{\overline{Y}}^{-1}(s;1)\frac{\partial f(x_v,v;s)}{\partial x}$$

and $\underline{\overline{Y}}^{-1}(1;1) = I_n$.

Therefore,

$$\frac{d\,\overline{Y}_n^{-1}(s;1)}{ds} = \overline{Y}_n^{-1}(s;1)\,\frac{\partial f}{\partial x}(x_v,v;s) .$$

Define $\lambda(t) \equiv \overline{Y}_n^{-1}(t;1)$, a row vector valued function. Then,

$$\lambda'(t) = -\lambda(t)\frac{\partial f}{\partial x}(x_v,v;t)$$

$$\lambda_i(1) = 0 \qquad i = 1,\ldots,n-1$$

$$\lambda_n(1) = 1 .$$

Since $H(x,u,\lambda;t) = f_1(x,u;t)\lambda_1 +\ldots+ f_n(x,u;t)\lambda_n$

$$\frac{\partial H}{\partial x}(t) = \left[\frac{\partial H}{\partial x_1},\ldots,\frac{\partial H}{\partial xn}\right]$$

$$= \lambda(t)\frac{\partial f}{\partial x}(x,u;t) .$$

Thus, $\lambda'(t) = -\frac{\partial H}{\partial x}(x,v,\lambda;t)\Big|_{x = x(v;t)}$

implies that (2), (3) and (4) of theorem 1.1 are satisfied. Note that,

$$\lambda^{T'}(t) = -\frac{\partial f^T}{\partial x}(x,v;t)\lambda^T(t) \quad \text{and}$$

$$\frac{\partial H}{\partial x}(x,u,\lambda;t) = \lambda(t)\frac{\partial f}{\partial u}(x,u;t) .$$

Equation (1.6) becomes, where $x = x(v;t)$

$$DJ(v)h = \int_0^1 \lambda(s)\frac{\partial f}{\partial u}(x,v;s)h(s)ds$$

$$= \int_0^1 \frac{\partial H}{\partial u}(x,v;s)h(s)ds$$

For all $h\epsilon\Omega_A - v$ by remark 1.5 part (2) since v is an optimal control, Ω_A is convex and the derivative exists,

$$DJ(v)h = \int_0^1 \frac{\partial H}{\partial u}(x,v;s)h(s)ds \leq 0.$$

Next it is shown that the integrand is less than or equal to zero.

Let $h(t) = w(t) - v(t)$ for any $w(t)\varepsilon\Omega_A$. Then the above inequality implies that

$$\int_0^1 \frac{\partial H}{\partial u}(x,v,\lambda;s)w(s)ds \leq \int_0^1 \frac{\partial H}{\partial u}(x,v,\lambda;s)v(s)ds \ . \qquad (1.7)$$

Suppose there exists $\overline{w}\varepsilon\Omega_A$ such that

$$\frac{\partial H}{\partial u}(x,v,\lambda;t)\overline{w}(t) > \frac{\partial H}{\partial u}(x,v,\lambda;t)v(t)$$

for $t\varepsilon C$ and $\mu(C)>0$.

Let $$w^*(t) = \begin{cases} v(t) & t\varepsilon[0,1]\setminus C \\ \overline{w}(t) & t\varepsilon C \end{cases}$$

By remark 1.4, $w^*(t)\varepsilon\Omega_A$. However, this implies that

$$\int_0^1 \frac{\partial H}{\partial u}(x,v,\lambda;s)w^*(s)ds > \int_0^1 \frac{\partial H}{\partial u}(x,v,\lambda;s)v(s)ds$$

which contradicts (1.7). Therefore

$$\frac{\partial H}{\partial u}(x,v,\lambda;s)h(s) \leq 0 \quad \text{a.e. for all } h\varepsilon\Omega_A - v. \qquad (1.8)$$

By assumption $f(x,u;t)$ is concave with respect to u in N_v which implies $H(x,u,\lambda;t) = f^T(x,u;t)\lambda^T(t)$ is concave with respect to u. Note that as a function of u, $K(u) = H(x_v,u,\lambda_v;t)$ has differential,

$$DK(u)\overline{h} = \frac{\partial H}{\partial u}(x_v,u,\lambda_v;t)\overline{h}(t)$$

for all $\overline{h}\varepsilon\Omega_A - u$. But H concave and (1.8) imply by remark 1.5 part (1) that v is also the optimal function for $H(x_v,u,\lambda;t)$. Therefore,

$$H(x_v,v,\lambda_v;t) \geq H(x_v,u,\lambda_v;t) \text{ for all } u\varepsilon\Omega_A \cap N_v$$

and the theorem is proved.

Note that if Ω_A is Ω_A^2, then

$$DJ(v)h = \int_0^1 \frac{\partial H}{\partial u}(x,v,\lambda;s)h(s) = 0 \ .$$

Since the above is true for all $h\varepsilon\Omega_A - v = \Omega_A^2 - v = \Omega_A^2$, choose $h(s) =$

sgn $\left(\frac{\partial H}{\partial u}(x,v,\lambda;s)\right)\varepsilon\Omega_A^2$ (see remark 3.4). This implies that

$$\frac{\partial H}{\partial u}(x,v,\lambda;s) = 0 \quad \text{a.e.} \quad \text{and}$$

$$\frac{\partial H}{\partial u}(x,v,\lambda;s)h(s) = 0 \quad \text{a.e.} \; .$$

Thus v is an interior maximum for H.

Definition 1.1 Let $f:S\subset\overline{X}\to\overline{Y}$, f differentiable. $x^*\varepsilon S$ is called a stationary point of f if

$$Df(x^*,h)\leq 0 \;, \text{ for all } h\varepsilon S-x^*.$$

Drop the concavity assumption for f(x,u;t) in theorem 1.1. Then theorem 1.1 says the following. If v maximizes the cost functional, then v is a stationary point for the Hamiltonian.

Remark 1.21 If f(x,u;t) is independent of x_n, then $\lambda_n(t) \equiv 1$ and $H(x,u,\lambda;t) = \overline{f}^T(\overline{x},u;t)\overline{\lambda}^T(t) + L(\overline{x},u;t)$

where $\overline{f} = (f_1,\ldots,f_{n-1})^T$, $\overline{x} = (x_1,\ldots,x_{n-1})^T$ and $\overline{\lambda} = (\lambda_1,\ldots,\lambda_{n-1})$

Proof:

$$\lambda'(t) = -\lambda(t)\frac{\partial f}{\partial x}(x,v;t)$$

$$\lambda_n'(t) = -\lambda(t)\left(\frac{\partial f}{\partial x}(x,v;t)\right) n^{th} \text{ column}$$

$$= -\lambda(t)\left(\frac{\partial f1}{\partial x_n},\ldots,\frac{\partial fn}{\partial x_n}\right)^T$$

$$= 0 \; .$$

Since $\lambda_n(1) = 1$, $\lambda_n(t) \equiv 1$.

Since $f_n(\overline{x},u;t) = L(\overline{x},u;t)$ and $\lambda_n(t) \equiv 1$

$$H(x,u,\lambda;t) = f^T(x,u;t)(\lambda_1(t),\ldots,\lambda_{n-1}(t),1)^T$$

$$= \overline{f}^T(\overline{x},u;t)\overline{\lambda}^T(t) + f_n(\overline{x},u;t)\cdot 1$$

$$= \overline{f}^T(\overline{x},u;t)\overline{\lambda}^T(t) + L(\overline{x},u;t)$$

which was to be shown.

The linear quadratic as defined (0.2) has the differential constraint independent of x_n. For notational convenience, let the dimension of the first n-1 differential constraints of the linear quadratic problem be n where problem (1.1) can be thought of having been n+1. Therefore, (0.2) is restated as follows.

$$\max_{u \varepsilon \Omega_A} J(u) = \tfrac{1}{2}\int_0^1 (x^T(t)Qx(t) + u^T(t)Ru(t))dt$$

$$x'(u;t) = Ax(u;t) + Bu(t)$$

$$x(u;o) = x_0$$

$x(t)\varepsilon R^n$, $u(t)\varepsilon R^m$, A nxn matrix, B nxm matrix,

Q negative semidefinite nxn matrix, and R negative definite mxm matrix.

By remark 1.21

$$H(x,u,\lambda;t) = (Ax(t) + Bu(t))^T\lambda(t) + \tfrac{1}{2}x^T(t)Qx(t) + \tfrac{1}{2}u^T(t)Ru(t)$$

$$\lambda'(t) = -\frac{\partial H(x,u,\lambda;t)}{\partial x} = -A^T\lambda(t) - Qx(t) \qquad (1.9)$$

$$\lambda(1) = 0$$

$\lambda(t)\varepsilon R^n$ where $\lambda(t)$ is considered a column vector and is $\lambda^T(t)$ of theorem 1.1 except for the last component.

Remark 1.22 The linear quadratic problem satisfies the conditions to theorem 1.1.

Proof $f(x,u;t) = Ax(u;t) + Bu(t)$. As a function of x, f is continuous for fixed u and as a function of u, f is continuous for fixed x. Since t does not occur explicitly; that is, $f(x,u;t) = f(x,u)$, f is continuous for all $t\varepsilon[0,1]$. Moreover, since f is affine in x and u, f is continuous in $(x(t),u(t))\varepsilon R^n x R^m$. A similar argument shows that

$$L(x,u;t) = \tfrac{1}{2}x^T(t)Qx(t) + \tfrac{1}{2}u^T(t)Ru(t)$$

is continuous in $(x(t),u(t);t)\varepsilon R^n x R^m x[0,1]$.

$$\frac{\partial f}{\partial x}(x,u;t) = A$$

$$\frac{\partial L}{\partial x}(x,u;t) = x^T(t)Q$$

$$\frac{\partial f}{\partial u}(x,u;t) = B$$

$$\frac{\partial L}{\partial u}(x,u;t) = u^T(t)R$$

All the above are continuous in $R^n x R^m x[0,1]$ and satisfy a Lipschitz condition in x and u.

$$||f(x,u;t)||_n \leq ||B||_s||u||_m + ||A||_s|x(u;t)||_n$$

Let $0\leq\alpha\leq 1$ and $\beta = (1-\alpha)$.

$$f(x,\alpha u_1+\beta u_2;t) = \alpha f(x,u_1;t) + \beta f(x,u_2;t) .$$

Thus, f is concave with respect to u.

$$L(x,\alpha u_1+\beta u_2;t) = \tfrac{1}{2}x^T(t)Qx(t) + \tfrac{1}{2}\alpha^2 u_1{}^T(t)Ru_1(t) +$$
$$\alpha\beta u_1{}^T(t)Ru_2(t) + \tfrac{1}{2}\beta^2 u_2{}^T(t)Ru_2(t) .$$

Therefore $L(x,\alpha u_1+\beta u_2;t) - (\alpha L(x,u_1;t) + \beta L(x,u_2;t)) =$

$$\tfrac{1}{2}(\alpha^2-\alpha)u_1{}^T(t)Ru_1(t) + \alpha\beta u_1{}^T(t)Ru_2(t) + \tfrac{1}{2}(\beta^2-\beta)u_2{}^T(t)Ru_2(t) =$$

$$\tfrac{1}{2}\alpha\beta((u_1(t)+u_2(t))^T R(u_1(t)+u_2(t)))<0 .$$

Thus, L(x,u;t) is strictly concave with respect to u. All the conditions for theorem 1.1 hold for all $u\varepsilon L_2^m[0,1]$ and the remark is proved.

1.2 Examples

Example 1 Let $u\varepsilon\Omega_A$, u twice differentiable.

$$x'(u;t) = u(t)$$

$$x(o) = 1$$

$$\max_{u\varepsilon\Omega_A} J(u) = -\tfrac{1}{2}\int_0^1 (x^2(t)+u^2(t))dt .$$

The problem is linear quadratic. Using remark 1.21 and the equations (1.9)

$$x(t) = 1 + \int_0^t u(t)dt$$

$$H(x,u,\lambda;t) = f(x,u;t) \cdot \lambda(t)$$

$$= u(t)\lambda(t) - \tfrac{1}{2}(x^2(t) + u^2(t))$$

$$\frac{\partial H}{\partial x}(x,u,\lambda;t) = -x(t)$$

$$\lambda'(t) = x(t)$$

$\lambda(1) = 0$ which implies that

$$\lambda(t) = - \int_t^1 x(s)ds$$

$$= -\int_t^1 (1+\int_0^s u(r)dr)ds$$

$$\frac{\partial H}{\partial u}(x,u,\lambda;t) = \lambda(t) - u(t)$$

$$= -\int_t^1 (1+\int_0^s u(r)dr)ds - u(t) .$$

From (1.6)

$$DJ(u)h = \int_0^1 \frac{\partial H}{\partial u}(x,u,\lambda;t)h(t)dt$$

$$= - \int_0^1 (\int_t^1 (1+\int_0^s u(r)dr)ds+u(t))h(t)dt .$$

For the unconstrained case, $DJ(v)h = 0$ for all $h\varepsilon\Omega_A - v$ which implies that $\frac{\partial H}{\partial u}(x,v,\lambda;t) = 0$.

Thus,

$$-\int_t^1 (1+\int_0^s v(r)dr)ds - v(t) = 0 \quad \text{or}$$

$$\int_t^1 \int_0^s v(r)drds + v(t) = t-1 \; . \qquad (1.10)$$

Hence, $v(1) = 0$. Differentiating (1.10),

$$-\int_0^t v(r)dr + v'(t) = 1 \qquad (1.11)$$

$v'(o) = 1$. Differentiating (1.11),

$v(t) - v''(t) = 0$ which implies

$$v(t) = ae^t + be^{-t}$$

$$v(1) = 0 = ae + be^{-1}$$

$$v'(t) = ae^t - be^{-t}$$

$$v'(o) = 1 = a - b$$

or $a = \frac{1}{e^2+1}$ $\quad b = -\frac{e^2}{1+e^2}$ and

$$v(t) = \frac{1}{e^2+1}(e^t - e^2 e^{-t})$$

$$v'(t) = \frac{1}{e^2+1}(e^{+} + e^2 e^{-t}) > 0 \quad \text{thus } |v(t)| \le \frac{e^2-1}{e^2+1} < 1 \; .$$

The unconstrained solution is also the constrained solution.

$$x(t) = 1 + \frac{1}{e^2+1}\int_0^t (e^s - e^2 e^{-s})dt$$

$$= \frac{1}{e^2+1}(e^t + e^2 e^{-t}) \; .$$

Example 2: Let $u \varepsilon \Omega_A = \Omega_A^1$

$$x'(t) = -x(t) + u(t)$$

$$x(o) = 1$$

$$\max_{u \varepsilon \Omega_A} J(u) = -\tfrac{1}{2}\int_0^1 u^2(t)dt \,.$$

The problem is linear quadratic. Using remark 1.21 and equations (1.9)

$$H(x,u,\lambda;t) = -x(t)\lambda(t) + u(t)\lambda(t) - \tfrac{1}{2}u^2(t)$$

$$\frac{\partial H}{\partial u}(x,u,\lambda;t) = \lambda(t)-u(t) \,.$$

Assume the problem is unconstrained. This implies that

$$DJ(v)h = o \quad \text{or that} \quad \frac{\partial H}{\partial u}(x,v,\lambda;t)\Big|_{x = x(v;t)} = 0 \,.$$

Thus, $\lambda(t) - v(t) = 0$ or

$$v(t) = \lambda(t)$$

$$\frac{\partial H}{\partial x}(x,v,\lambda;t) = -\lambda(t)$$

$$\lambda'(t) = \lambda(t)$$

$\lambda(t) = c \cdot e^t$ but $\lambda(1) = 0$ means that $c = 0$.

Therefore, $\lambda(t) \equiv 0$ so that $v(t) = 0$. Now

$$x'(t) = -x(t)$$

$x(0) = 1$ implies that

$$x(t) = e^{-t} \,.$$

Note that for the initial valued differential constraint problem where $\overline{f}(\overline{x},u;t)$ is a function of $x_1,\ldots,x_{n-1}$, u and t with a cost functional integrand $L(\bar{x},u;t) = u^{2k}(t)$

$$v(t) \equiv 0 \ .$$

In many applications, the differential constraint is a boundary valued problem. The optimal control in this case is not identically zero in most cases.

2. UNCONSTRAINED OPTIMAL CONTROL PROBLEMS AND INTEGRAL EQUATIONS

2.1 Optimal Control Problems as Integral Equations

This chapter deals with locally normal (see definition 0.1) unconstrained optimal control problems which satisfy the conditions to theorem 1.1. Call this class of optimal control problems, class $\overline{Q}$. Class $\overline{Q}$ problems will be solved by transforming the original problem into an integral equation which will then be put into a form suitable for finding approximate solutions. The linear quadratic problem will be shown to be in class $\overline{Q}$. It is transformed into a linear integral equation and it will be shown that the kernel is continuous and the first derivative of the kernel with respect to the integrating factor is bounded on [0,1]. Thus, guaranteed error bounds for the approximate solution of the state function will be obtained for linear quadratic problem using interval analytic techniques developed by Winter (17) and is discussed in chapter 4 and appendix A.

Nonlinear class $\overline{Q}$ problems can be transformed into nonlinear integral equations with a kernel that is discontinuous with respect to its integrating factor. Since Winter (17) assumes the kernel of the nonlinear integral equation is continuous with respect to the integrating factor, his method does not apply. Approximate solutions to the nonlinear problem can be obtained, and in chapter 4 methods for obtaining guaranteed bounds are suggested. Before discussing linear quadratic problems, the conditions necessary for an optimal control problem to be locally normal are developed. First, the implicit function theorem is stated without proof.

Implicit Function Theorem (Dieudonné (5)) Let E,F,G be three Banach spaces, f continuously differentiable mapping of an open subset A of ExF into G. Let (x_0,y_0) be a point of A such that $f(x_0,y_0) = 0$ and the partial derivative $\frac{\partial f}{\partial y}(x_0,y_0)$ be invertible where $\frac{\partial f}{\partial y}(x_0,y_0)$ is a linear mapping of F onto G. Then there is an open neighborhood U_0 of x_0 in E such that for every connected neighborhood U of x_0 contained in U_0, there is a unique continuous mapping u of U into F such that $u(x_0) = y_0$, $(x,u(x))\varepsilon A$ and $f(x,u(x)) = 0$ for any $x\varepsilon U$. Furthermore, u is continuously differentiable and its derivative is given by

$$u'(x) = -\left(\frac{\partial f}{\partial y}(x,u(x))\right)^{-1}\cdot\left(\frac{\partial f}{\partial x}(x,u(x))\right).$$

It has been seen in the discussion following theorem 1.1 that the unconstrained problem can be reduced to solving,

$$\frac{\partial H}{\partial u}(v) = 0 \quad \text{a.e.}\ .$$

If $\frac{\partial H}{\partial u}$ satisfies the conditions of the implicit function theorem where $(x_v,\lambda_v;t) = x_0$ in the above, then the unconstrained problem is locally normal. For the linear quadratic problem,

$$\frac{\partial H}{\partial u}(u) = \lambda^T(t)B + u^T(t)R$$

which in continuously differentiable in $(x(t),\lambda(t),t)$ and,

$$\frac{\partial^2 H}{\partial u^2}(x,u,\lambda;t) = R$$

which is invertible. Therefore, linear quadratic problems are normal which could have been determined by inspection.

The conditions to theorem 1.1 are not enough to imply that the optimal control problem (1.1) where $u\varepsilon\Omega_A^{\cdot\cdot}$ is locally normal.

Remark 2.1 Let the conditions to theorem 1.1 hold. Assume that $\frac{\partial^2 f^i}{\partial u^2}$ and $\frac{\partial^2 f^i}{\partial x \partial u}$ exist and are continuous in a neighborhood N_v^{i*} of $(x_v(t),v(t),\lambda_v(t),t^i)$ for $t^i\varepsilon(t_{i-1},t_i)$, $i=1\ldots,N$. Assume

$\lambda(t)\frac{\partial^2 f^i}{\partial u^2}(x,u;t^i)$ is invertible in N_v^{i*}. Then the optimal control problem (1.1) is locally normal.

Proof Keeping the notation from theorem 1.1

$$\frac{\partial H}{\partial u}(u) = \lambda(t)\frac{\partial f^i}{\partial u}(x,u;t). \quad \text{Thus,}$$

$$\frac{\partial^2 H}{\partial x \partial u}(u) = \lambda(t)\frac{\partial^2 f^i}{\partial x \partial u}(x,u;t)$$

$$\frac{\partial^2 H}{\partial \lambda \partial u}(u) = \frac{\partial f^i}{\partial u}(x,u;t)$$

$$\frac{\partial^2 H}{\partial u^2}(u) = \lambda(t)\frac{\partial^2 f^i}{\partial u^2}(x,u;t) .$$

The application of theorem 1.1 to unconstrained optimal control problems implies that

$$\frac{\partial H}{\partial u}(v) = 0 \quad \text{a.e.} ,$$

The conditions of remark 2.1 imply that the implicit function theorem can be applied where $v(t^1)=y_0$ and $(x(t')\lambda(t^1),t^1)$ is x_0 in the theorem. $i = 1...,N$. Thus the remark is proved.

Definition 2.1 An optimal control problem is in class Q if it satisfies the conditions of remark 2.1.

Class Q is nonempty since it contains the linear quadratic problems. Moreover, unless an explicit function $q^i(x,\lambda;t)$ is found for which $v(t) = q^i(x,\lambda;t)$ where $\frac{\partial H}{\partial u}(x_v,q^i(x,\lambda;t),\lambda;t) = 0$ and $q^i(x,\lambda;t)$ exists for $t\varepsilon(t_{i-1},t_i)$ $i = 1,...,N$, the method is theoretically useful, but in practice, useless.

2.1.1 Unconstrained Linear Quadratic Problems - The linear quadratic problem is given by (1.9). Since as a column vector,

$$\frac{\partial H^T}{\partial u}(v) = B^T\lambda(t) + Rv(t) = 0 \qquad \text{and}$$

$$v(t) = - R^{-1}B^T\lambda(t) . \qquad (2.1)$$

Substituting for u(t) in (1.9)

$$x'(t) = Ax(t) - BR^{-1}B^T\lambda(t)$$

$$\lambda'(t) = -A^T\lambda(t) - Qx(t)$$

$$x(t) = e^{At}x_0 - \int_0^t e^{A(t-s)}BR^{-1}B^T\lambda(s)ds$$

$$\lambda(t) = \int_t^1 e^{-A^T(t-s)}Qx(s)ds \qquad (2.2)$$

or

$$x(t) = e^{At}x_0 + \int_0^1 G_1(s,t)\lambda(s)ds \qquad (2.3)$$

where

$$G_1(s,t) = \begin{cases} -e^{A(t-s)}BR^{-1}B^T & 0\le s\le t\le 1 \\ 0 & 0\le t<s\le 1 \end{cases}$$

$$\lambda(t) = \int_0^1 G_2(s,t)x(s)ds \qquad (2.4)$$

where

$$G_2(s,t) = \begin{cases} 0 & 0<s<t\le 1 \\ e^{-A^T(t-s)}Q & 0\le t\le s\le 1 \ . \end{cases}$$

Substituting (2.4) into (2.3)

$$x(t) = e^{At}x_0 + \int_0^1 \left(\int_0^1 G_1(s,t)G_2(r,s)ds\right)x(r)dr$$

For $0\le r\le t\le 1$, let

$$k_1(r,t) = \int_0^1 G_1(s,t)G_2(r,s)ds$$

$$= -\int_0^r e^{A(t-s)}BR^{-1}B^Te^{-A^T(s-r)}Qds \ . \qquad (2.5)$$

For $0\le t<r\le 1$, let

$$k_2(r,t) = \int_0^1 G_1(s,t)G_2(r,s)ds$$

$$= -\int_0^t e^{A(t-s)}BR^{-1}B^Te^{-A^T(s-r)}Qds \ . \qquad (2.6)$$

From the above $k_1(r,r) = k_2(r,r)$.

$$\text{Let } k(r,t) = \begin{cases} k_1(r,t) & o \leq r \leq t \leq 1 \\ k_2(r,t) & o \leq t < r \leq 1 \end{cases} \tag{2.7}$$

The function k(r,t) is continuous for $o \leq r,t \leq 1$. The state function is the linear integral equation

$$x(t) = e^{At}x_0 + \int_0^1 k(r,t)x(r)dr \tag{2.8}$$

where k(r,t) is defined by (2.7). Let $D \equiv BR^{-1}B^T$

$$\frac{\partial k_1(r,t)}{\partial r} = -\, e^{A(t-r)}DQ - e^{At}\left(\int_0^t e^{-As}De^{-A^Ts}ds\right)A^Te^{A^Tr}Q$$

$$\frac{\partial k_2(r,t)}{\partial r} = -\, e^{At}\left(\int_0^t e^{-As}De^{-A^Ts}ds\right)A^Te^{A^Tr}Q$$

$\frac{\partial k_1}{\partial r}$ and $\frac{\partial k_2}{\partial r}$ are continuous for $o \leq r,\ t \leq 1$, but $\frac{\partial k_1(r,r)}{\partial r} - \frac{\partial k_2(r,r)}{\partial r} = -\,DQ$.

Therefore, $\frac{\partial k(r,t)}{\partial r}$ is discontinuous at $r = t$. However, it can be seen that

$$\left|\left|\frac{\partial k}{\partial r}\right|\right|_M \leq L \qquad \text{on } [0,1]\,.$$

Equation (2.8) and the above inequality are used in conjunction with interval analytic techniques to obtain approximate solutions to the linear quadratic problem with guaranteed error bounds. The method is described in chapter 4 and appendix A.

2.1.2 Nonlinear Problems - Nonlinear optimal control problems often have no explicit formula for v in terms of x_v and λ_v as do all linear quadratic problems. Assume the nonlinear problem belongs to class Q. This means by the implicit function theorem, that there exist continuous functions $q^i(x,\lambda;t)$, $i = 1...,N$ such that $q^i(x,\lambda;t)$ is differentiable in N_v^{i*} and

$$v(t) = q^i(x,\lambda;t) \qquad t\varepsilon(t_{i-1},t_i)\,. \tag{2.9}$$

Equation (2.9) is substituted into the differential equations describing x and λ where $q(x,\lambda;t) = q^i(x,\lambda;t)$ for $t\varepsilon(t_{i-1},t_i)$. Here $q:R^n X R^m x O(f) \to R^n$ where it is understood that $q(x,\lambda;t) = q(x(t),\lambda(t);t)$. Let $t\varepsilon O(f)$ and $N_v^* = \bigcup_{i=1}^{N} N_v^{i*}$. Then

$$x'(t) = f(x,q(x,\lambda;t),t)$$

$$= d(x,\lambda;t) \qquad (2.10)$$

$$x(o) = x_o$$

$$\lambda'(t) = - \frac{\partial H}{\partial x}(x,q(x,\lambda;t),\lambda;t)$$

$$= - \lambda(t)\frac{\partial f}{\partial x}(x,q(x,\lambda;t);t) \qquad (2.11)$$

$$= e(x,\lambda;t)$$

$$\lambda(1) = (0,\ldots,1) .$$

Recall $\lambda(t)$ is a row vector. Define

$$y(t) = (x^T(t),\lambda(t))^T,$$

$$y'(t) = (d^T(y;t),e(y;t))^T$$

$$= F(y;t) \quad \text{and}$$

$$c = (x_1(o),\ldots,x_n(o),\lambda_1(1),\ldots,\lambda_n(1))^T .$$

Next the solution to (2.10) and (2.11) is written as a nonlinear integral equation. The method used to obtain the integral equation in this thesis greatly simplifies the method that Falb and deJong (6) use to transform nonlinear unconstrained optimal control problems into integral equations.

Let,

$$y(t) = c + \int_0^1 k(s,t)F(y;s)ds \quad \text{and}$$

$$k_1(s,t) = \begin{pmatrix} I_n & 0_n \\ 0_n & 0_n \end{pmatrix} \quad 0 \le s \le t \le 1 \tag{2.12}$$

$$k_2(s,t) = \begin{pmatrix} 0_n & 0_n \\ 0_n & -I_n \end{pmatrix} \quad 0 \le t < s \le 1 . \tag{2.13}$$

Define

$$k(s,t) = \begin{cases} k_1(s,t) & 0 \le s \le t \le 1 \\ k_2(s,t) & 0 \le t < s \le 1 . \end{cases} \tag{2.14}$$

It must be verified that with the above kernel (2.14), the two endpoint conditions are satisfied and that the derivative of $y(t)$ is $F(y;t)$, where $y(t)$ is described by the above integral equation.

$$y(o) = c + \int_0^1 k_2(s,o)F(y;s)ds$$

$$= c + \int_0^1 \begin{pmatrix} 0_n & 0_n \\ 0_n & -I_n \end{pmatrix} \begin{pmatrix} d(y;s) \\ e^T(y;s) \end{pmatrix} ds$$

$$= c - \int_0^1 \begin{pmatrix} 0 \\ e^T(y;s) \end{pmatrix} ds$$

$$= c + (0,\dots,0,\lambda_1(o)-\lambda_1(1),\dots,\lambda_n(o)-\lambda_n(1))^T .$$

Thus, the left endpoint condition is satisfied.

$$y(1) = c + \int_0^1 k_1(s,1)F(y;s)ds$$

$$= c + \int_0^1 \begin{pmatrix} I_n & 0_n \\ 0_n & 0_n \end{pmatrix} \begin{pmatrix} d(y;s) \\ e^T(y;s) \end{pmatrix} ds$$

$$= c + \int_0^1 \begin{pmatrix} d(y;s) \\ 0 \end{pmatrix}$$

$$= c + (x_1(1)-x_1(o),\dots,x_n(1)-x_n(o),0,\dots,0)^T .$$

Thus the right endpoint condition is satisfied.

$$y'(t) = \frac{\partial}{\partial t}\left(\int_0^t k_1(s,t)F(y;s)ds\right) + \frac{\partial}{\partial t}\left(\int_t^1 k_2(s,t)F(y;s)ds\right)$$

$$= k_1(t;t)F(y;t) - k_2(t;t)F(y;t)$$

$$= F(y;t) .$$

Thus, the derivative condition is satisfied. And

$$y(t) = c + \int_0^1 k(s,t)F(y;s)ds \qquad (2.15)$$

is an integral equation representation for the nonlinear optimal control problem. Clearly the kernel of (2.15) is discontinuous at $0 \leq s = t \leq 1$.

<u>Remark 2.2</u> $F(y;t)$ is continuous in y and piecewise continuous in t in N_V^* for class Q optimal control problems.

<u>Proof</u> From (2.10) and (2.11),

$$F(y;t) = \begin{pmatrix} f(x,q(x,\lambda;t);t) \\ -\frac{\partial f}{\partial x}^T(x,q(x,\lambda;t);t)\lambda^T(t) \end{pmatrix} .$$

The function f is continuous in $R^n x R^m x O(f)$ by assumption (i) of theorem 1.1 and from the implicit function theorem, $q(x,\lambda;t)$ is continuous in x,λ and piecewise continuous in t in N_V^*. Moreover, by assumption (ii) to theorem 1.1, $\frac{\partial f}{\partial x}$ is continuous in $R^n x R^m x O(f)$ while $\lambda(t)$ is continuous in t. Therefore, the composition $f(x,q(x,\lambda;t);t)$ is continuous in x, and piecewise continuous in t. Likewise, the composition and product $\lambda(t)\frac{\partial f}{\partial x}(x,q(x,\lambda;t);t)$ is continuous in x, and piecewise continuous in t. Thus the remark is proved.

Numerical methods for obtaining approximate solutions to (2.15) are discussed in chapter 4.

2.2 Examples

Example 1 Let $u\varepsilon\Omega_A$ unconstrained.

$$x'(t) = u(t)$$

$$x(o) = x_o$$

$$J(u) = -\tfrac{1}{2}\int_0^1 (x^2(t) + u^2(t))dt .$$

This example is example 1 of section 1.2.

$$\lambda'(t) = x(t) \qquad , \quad \lambda(1) = 0$$

$$v(t) = \lambda(t)$$

and

$$x'(t) = \lambda(t)$$

$$\lambda'(t) = x(t) .$$

Substituting (2.5) and (2.6) where A = 0, B = 1, Q = -1, R = -1,

$$k(r,t) + \begin{cases} k_1(r,t) = -r & o \leq r \leq t \leq 1 \\ k_2(r,t) = -t & o \leq t < r \leq 1 . \end{cases}$$

Example 2 Let $u\varepsilon\Omega_A$ unconstrained.

$$x'(t) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix}$$

and

$$J(u) = \tfrac{1}{2}\int_0^1 (x^T(t)\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} x(t) + u^T(t)\begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix} u(t))dt$$

$$\lambda'(t) = -A^T\lambda(t) - Qx(t)$$

$$= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \lambda(t) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} x(t)$$

$$v(t) = -R^{-1}B^{T}\lambda(t)$$

$$= \tfrac{1}{2}\begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \lambda(t)$$

$$e^{At} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}$$

$$-A^{T} = A$$

$$e^{A^{T}t} = (e^{At})^{T} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

$D = BR^{-1}B^{T} = -\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Using (2.5) and (2.6)

$$k_1(r,t) = -\begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \int_0^r \begin{pmatrix} \cos(s) & \sin(s) \\ -\sin(s) & \cos(s) \end{pmatrix} \begin{pmatrix} \cos(s) & -\sin(s) \\ \sin(s) & \cos(s) \end{pmatrix} ds X$$

$$\begin{pmatrix} \cos r & \sin r \\ -\sin r & \cos r \end{pmatrix}$$

$$= -r \begin{pmatrix} \cos(r+t) & -\sin(r+t) \\ \sin(r+t) & \cos(r+t) \end{pmatrix}.$$ Likewise,

$$k_2(r,t) = -t \begin{pmatrix} \cos(r+t) & -\sin(r+t) \\ \sin(r+t) & \cos(r+t) \end{pmatrix}$$

Example 3 Let $u \in \Omega_A$ unconstrained.

$$x'(t) = -x^2(t) + u(t)$$

$$x(0) = x_0$$

$$J(u) = -\tfrac{1}{2} \int_0^1 (x^2(t) + u^2(t))dt$$

$$H(x,u,\lambda;t) = -x^2(t)\lambda(t) + u(t)\lambda(t) - \tfrac{1}{2}x^2(t) - \tfrac{1}{2}u^2(t)$$

$$\frac{\partial H}{\partial u}_u(t) = \lambda(t) - u(t)$$

$$\lambda'(t) = -\frac{\partial H}{\partial x}_u(t) = 2x(t)\lambda(t) + x(t)$$

$$v(t) = \lambda(t)$$

$$y'(t) = \begin{pmatrix} -x^2(t) + \lambda(t) \\ 2x(t)\lambda(t) + x(t) \end{pmatrix} = F(y(t);t) .$$

Using (2.15) $k(s,t)F(y,s) = \begin{cases} \begin{bmatrix} -x^2(s) + \lambda(s) \\ 0 \end{bmatrix} & 0 \le s < t \le 1 \\ \begin{bmatrix} 0 \\ 2x(s)\lambda(s) + x(s) \end{bmatrix} & 0 \le t < s \le 1 \end{cases}$

and $\quad y(t) = (x_0,0)^T + \int_0^1 k(s,t)F(y;s)ds .$

3. CONSTRAINED OPTIMAL CONTROL PROBLEMS AND THE CONDITIONAL GRADIENT ALGORITHM

3.1 Introduction

This chapter applies the conditional gradient method to constrained optimal control problems.

Definition 3.1 Let $f: S \subset \underline{X} \to R$ be a functional from the subset S of a normed linear space $\underline{X}$ into the real numbers. Suppose there is $x^* \varepsilon S$ such that

$$f(x^*) \geq f(x) \text{ for all } x \varepsilon S.$$

Then f will be called locally concave if f(x) is concave in $N_{x^*} = \{x : x \varepsilon S, ||x-x^*||_{\underline{X}} \leq \varepsilon, \varepsilon > o\}$. The class of optimal control problems on which the conditional gradient method is applied will be those which satisfy the conditions of theorem 1.1 whose cost functionals are locally concave; call this class of optimal control problems class R. Unconstrained problems which are not in class Q can be solved by the conditional gradient method if they belong to class R. Furthermore, class R is not empty since it will be shown that the linear quadratic problems belong to class R.

The conditional gradient algorithm generates a sequence of controls $\{u_k\}$ if a function $u_1(t)$ can be found such that $u_1 \varepsilon \Omega_A$. If the differential constraints of (1.1) had included right hand endpoint conditions for x; that is, $x(u;1) = s_i$ for $1 \leq i \leq n-2$, finding an admissible initial control to start the conditional gradient becomes a problem in itself. For Ω_A as defined in theorem 1.1, $u_1(t) \equiv \overline{o}_m \varepsilon \Omega_A$. Thus, the conditional gradient method can always be started.

3.2 The Conditional Gradient Algorithm and its Application to Optimal Control Problems

The conditional gradient is often called the steepest descent method. The name conditional gradient has been used in conjunction with the application of the steepest descent method to optimal control problems by Denyanov and Rubinov (4) and Krabs (12). In this thesis, the name conditional gradient is used in conjunction with the steepest ascent method applied to optimal control problems.

Assume the optimal control problem belongs to class R and $u_k \varepsilon \Omega_A$ where u_k is in the convex neighborhood of concavity. A direction $h_k \varepsilon \Omega_A - u_k$ is chosen so that

$$DJ(u_k)h_k = \sup_{h\varepsilon\Omega_A - u_k} DJ(u_k)h \ .$$

It will be shown below that there exists such a h_k. If $DJ(u_k)h_k \leq o$ by remark 1.5, $v = u_k$ and the algorithm is terminated where $u_n = u_k$ for all $n \geq k$. Otherwise, $DJ(u_k)h_k > o$.

The step size $\alpha_k \varepsilon (0,1]$ is calculated as follows. The function

$$g_k(\alpha) \equiv J(u_k + \alpha h_k), \qquad 0 \leq \alpha \leq 1 \qquad (3.1)$$

is continuous, real valued and differentiable. Since $g'_k(0) = DJ(u_k)h_k > 0, g(\alpha) > g(0)$ in a righthand neighborhood of zero. Thus, let $\alpha_k \varepsilon (0,1]$ be a point for which $g_k(\alpha)$ is maximum. This implies

$$J(u_k + \alpha_k h_k) > J(u_k).$$

Let $\hat{u}_k = u_k + h_k$, so that $\hat{u}_k \varepsilon \Omega_A$

$$u_k + \alpha_k h_k = (1-\alpha_k)u_k + \alpha_k \hat{u}_k \varepsilon \Omega_A$$

since Ω_A is convex. Let $u_{k+1} = u_k + \alpha_k h_k$, replace u_k by u_{k+1} and restart the process.

A few lammata will be presented next. First the conditional gradient algorithm is restated in a general setting.

Let $f: S\subset\underline{X}\to R$, S convex and f concave. The conditional gradient algorithm is as follows:

1. Suppose $x_k \varepsilon S$

2. Pick $h_k \varepsilon S-x_k$ so that

   $Df(x_k, h_k) \geq Df(x_k, h)$ for all $h \varepsilon S-x_k$

3. If $Df(x_k, h_k) \leq 0$ set $x_n = x_k$

   for all $n>k$ and stop. Otherwise,

4. Pick $\alpha_k \varepsilon (0,1]$ so that

   $f(x_k + \alpha_k h_k) \geq f(x_k + \alpha h_k) \quad 0 \leq \alpha \leq 1$

5. Set $x_{k+1} = x_k + \alpha_k h_k$, replace k by k+1

   and go to step 2.

The conditional gradient method as applied to class R can be found in Demyanov and Rubinov (4) as well as in Krabs (12). The algorithm is developed here with explicit formulas to be used in conjunction with interval analytic techniques set forth in chapter 4. Moreover, if $D^2J(u)$ is negative definite, a bound on $||v-u_k||$ is obtained. It will be shown that for the linear quadratic problems, $D^2J(u)$ is negative definite which means that an interval extension function for v(t) can be obtained. For class R, J(v) will be shown to lie in an interval. Repeated use is made of the fact that continuous concave functions are weakly upper semicontinuous and that a weakly upper semicontinuous function attains its supremum on weakly compact sets.

Lemma 3.1 (Denyanov and Rubinov (4) or Krabs (12)) Let $\overline{\underline{X}}$ be a linear space and $f:\overline{\underline{X}}\to R$ concave. Then $Df(x,h)$ exists and $Df(x,h)$ is concave with respect to h for fixed x.

The proof of the above is elementary since the differential quotient

$$\frac{f(x+th) - f(x)}{t}$$

is nonincreasing.

Lemma 3.2 (Krabs (12)) Let $\overline{X}$ be a Banach space, $S \subset \overline{X}$ convex and weakly compact. Let $f: S \to R$, f continuous and concave. Assume $x_1 \varepsilon S$ and that there is $h_k \varepsilon S - x_k$ such that $Df(x_k, h_k) \geq Df(x_k, h)$ $h \varepsilon S - x_k$ where $\{x_k\}$ $k = 2, \ldots$ is a nonterminating conditional gradient sequence. If

$$Df(x_k, h_k) \to 0,$$

then

$$f(x_k) \to \sup_{x \varepsilon S} f(x)$$

Proof Let $f(z_k) \to \sup_{x \varepsilon S} f(x)$. By concavity and lemma 3.1,

$$f(z_k) - f(x_k) \leq Df(x_k, z_k - x_k) \, . \qquad (3.2)$$

By the definition of h_k,

$$f(z_k) - f(x_k) \leq Df(x_k, z_k - x_k) \leq Df(x_k, h_k)$$

and

$$\lim_{k \to \infty} f(z_k) \leq \lim_{k \to \infty} Df(x_k, h_k) + \lim_{k \to \infty} f(x_k) \, .$$

Therefore $\sup_{x \varepsilon S} f(x) \leq \lim_{k \to \infty} f(x_k) \leq \sup_{x \varepsilon S} f(x)$

which was to be shown.

Suppose that $\sup_{x \varepsilon S} f(x) = f(z)$, $z \varepsilon S$. By choosing $z_k = z$ for all k in lemma 3.2

$$Df(x_k, h_k) \geq f(z) - f(x_k) \geq 0 \qquad (3.3)$$

and

$$\lim_{k \to \infty} f(x_k) = f(z) \, .$$

Equation (3.3) gives a bound for $|f(z) - f(x_n)|$ which will be used in chapter 4.

Lemma 3.3 (Demyanov and Rubinov (4)) Let f, S, h_k and $\{x_k\}$ satisfy the conditions of lemma 3.2. In addition suppose that Df(x,h) is Lipschitz as a map from S to the space of bounded linear functionals from $\overline{X}$ to R. Then,

$$\lim_{n\to\infty} Df(x_n,h_n)\to 0 \quad .$$

Proof: Remark 1.7 with $||\cdot||_z = |\ |$, the absolute value where $||\ || = ||\ ||_{\overline{X}}$

$$|f(\overline{x}_{n+1}) - (f(x_n) + Df(x_n,\overline{x}_{n+1}-x_n))| \leq L_s||\overline{x}_{n+1} - x_n||^2 \quad .$$

Let $\overline{x}_{n+1} = x_n + \alpha h_n$ for any $\alpha\varepsilon(0,1)$. Substituting in the above

$$f(\overline{x}_{n+1}) - f(x_n) - \alpha Df(x_n,h_n) \geq -\alpha^2 L_s||h_n||^2$$

S convex, weakly compact implies S is bounded. Let $||h_n||^2 \leq M_s$. Moreover, $f(x_{n+1}) \geq f(\overline{x}_{n+1})$ where $x_{n+1} = x_n + \alpha_n h_n$. Thus,

$$\frac{f(x_{n+1}) - f(x_n)}{\alpha} + \alpha L_s M_s^2 \geq Df(x_n,h_n) > 0 \quad .$$

Let $\varepsilon>0$ be given and choose $\alpha = \frac{\varepsilon}{L_s M_s^2}$. Since S is weakly compact, there exists N such that for all $n\geq N$, $f(x_{n+1}) - f(x_n) \leq \varepsilon^2/4L_s M_s^2$. Then $0<Df(x_n,h_n)\leq\varepsilon$ which was to be shown.

The remarks and lemmata of chapter 1 prove that under the assumptions to theorem 1.1, DJ(u)h is Fréchet and satisfies a Lipschitz condition in $u\varepsilon\Omega_A$. For class R, DJ(u)h is concave with respect to h by lemma 3.1. The Fréchet derivative is continuous with respect to h and $\Omega_A'-u$ bounded. Therefore,

$\sup_{h\epsilon\Omega_A^{'}-u_k} DJ(u_k)h$ exists for some h in $\Omega_A^{'}-u_k$ and

$\lim_{n\to\infty} DJ(u_k)h_k \to 0$ by lemma 3.3 .

By lemma 3.2, $\lim_{k\to\infty} J(u_k) \to J(v)$. Equation (3.3) becomes

$$0 \leq J(v) - J(u_k) \leq DJ(u_k)h_k \quad . \qquad (3.4)$$

The function $h_k(t)$ is calculated as follows for class R and $\Omega_A = \Omega_A^{'}$

$$\sup DJ(u_k)h = \sup_{h\epsilon\Omega_A^{'}-u_k} \int_0^1 \frac{\partial H}{\partial u}^k(t)h(t)dt$$

$$= \sup_{h\epsilon\Omega_A^{'}-u_k} \left(\int_0^1 \frac{\partial H}{\partial u}^k(t)w(t)dt - \int_0^1 \frac{H}{u}^k(t)u_k(t)dt\right)$$

where $h = w - u_k$ for some $w\epsilon\Omega_A^{'}$. The second integral is constant. Thus, the supremum is taken over the first integral where the supremum of the left hand side of the above is achieved if

$$\sup_{w\epsilon\Omega_A^{'}} \int_0^1 \frac{\partial H}{\partial u}^k(t)w(t)dt$$

is achieved. Moreover, since

$$\int_0^1 \frac{\partial H}{\partial u}^k(t)w(t) \leq \int_0^1 \left|\frac{\partial H}{\partial u}^k(t)\right| \cdot |w(t)|dt$$

$$\leq \sum_{j=1}^{m} \int_0^1 \left|\frac{\partial H}{\partial u_j}^k(t)\right|dt$$

the supremum will be attained if there exists $w\epsilon\Omega_A^{'}$ such that the above inequality is an equality. If the j^{th} component of w is

$$w_j(t) \equiv \operatorname{sgn} \frac{\partial H}{\partial u_j}{}_k(t) \quad 1 \le j \le m$$

$$\equiv \begin{cases} 1 & \text{for } t \text{ such that } \frac{\partial H}{\partial u_j}{}_k(t) \ge 0 \\ -1 & \text{for } t \text{ such that } \frac{\partial H}{\partial u_j}{}_k(t) < 0 . \end{cases} \qquad (3.5)$$

then the above inequality is an equality. However, it must be shown that $w(t)$ as defined above belongs to Ω_A^1.

Remark 3.4 If $f(t)$ is a real valued function, $f \varepsilon L_2^1[0,1]$, then $\operatorname{sgn} f(t) \varepsilon L_2^1[0,1]$.

Proof: The range of sgn $f(t)$ is $(\{1\},\{-1\}),\{1\},\{-1\}$

$$(\operatorname{sgn} f)^{-1}(1) = f^{-1}([0.\infty)), \quad \text{and}$$

$(\operatorname{sgn} f)^{-1}(-1) = f^{-1}((-\infty .0))$. But f is given to be μ-measurable on $(0,\infty)$ and $(-\infty ,0)$ so that each of the above sets described by the two inverses are μ-measurable. Therefore, their unions are and sgn $f(t)$ is Lebesque integrable on $[0,1]$. Since $|\operatorname{sgn} f(t)| \le 1$, $\operatorname{sgn} f(t) \varepsilon L_2^1[0,1]$ which was to be shown.

Remark 3.4 implies that $w_j(t) \varepsilon L_2^1[0,1]$ and $w(t) \varepsilon L_2^m[0,1]$. $||w||_\infty \le 1$ by construction. Therefore, $w \varepsilon \Omega_A^1$.

Define $w(t) = (\operatorname{sgn} \frac{\partial H_k}{\partial u}(t))^T$ where $w_j(t)$ is defined by (3.5)

$$h_k(t) = \operatorname{sgn} \frac{\partial H}{\partial u}{}_k(t) - u_k(t) \qquad (3.6)$$

$$h_k \varepsilon \Omega_A^1 - u_k .$$

Note that if $\frac{\partial H}{\partial u}{}_k(t)=0$ a.e. then $DJ(u_k)h=0$ for all $h \varepsilon \Omega_A^1 - u_k$, u_k is the optimal control and the conditional gradient algorithm is stopped. In this case $h_k(t)$ can be any vector in $\Omega_A^1 - u_k$.

The conditional gradient algorithm in the unconstrained case would proceed as follows. It is assumed that there is an optimal control $v\varepsilon\Omega_A^2$. Let $\Omega_A = \Omega_n = \{u:u\ L_2^m[0,1]\ ,\ ||u||_\infty \leq n,\ n = 1, 2,\ldots\}$. For each n, the conditional gradient algorithm can be carried out as outlined above. Since $v\varepsilon\Omega_N$ for some finite N, the method produces an optimal cost functional in a finite number of steps. For the remaining portion of chapter 3, Ω_A is Ω_A' or Ω_n.

Remark 3.5 Let the optimal control problem belong to class R. Then

$$S_k \equiv \{u:u,u_k\varepsilon\Omega_A\ ,\ J(u)\geq J(u_k)\} \text{ is closed and convex.}$$

Proof By assumption J(u) is concave and continuous. Therefore S_k is closed. The convexity of S_k follows from the concavity of J(u) and the remark is proved.

If $\{u_k\}$ is the conditional gradient sequence, then

$$v\varepsilon S_{k+1}\subset S_k\subset \cdots \subset S_1 \subset \Omega_A \text{ for all } k\geq 1.$$

Lemma 3.6 Assume that for all $z(t) \neq \overline{0}_m$ a.e.

$$M||z||_m^2 \leq z^T D^2 J(u) z \leq m||z||_m^2 < 0$$

$u\varepsilon S_1$. Then for $\{u_k\}$ the conditional gradient sequence,

$$||v - u_k||_2 \leq \frac{2}{|m|}||DJ(u_k)|| \ . \tag{3.7}$$

Note that the hypothesis implies that J(u) is concave in S_1.

Proof: J(u) is a real valued function and its Taylor series is

$$\begin{aligned} J(u) - J(u_k) - DJ(u_k)(u-u_k) &= \tfrac{1}{2}(u-u_k)^T D^2 J(\overline{u})(u-u_k) \\ &\leq \tfrac{1}{2}m||u - u_k||_2^2 \leq 0 \end{aligned}$$

where $\overline{u} = \alpha u(t) + (1-\alpha)u_k(t) \quad \alpha\varepsilon(0,1),\ u\varepsilon S_k$.

Thus, $\overline{u}\varepsilon S_k$ by the convexity of S_k.

Therefore, since $J(u) \geq J(u_k)$ and

$$DJ(u_k)(u-u_k) \geq J(u) - J(u_k) \geq 0$$

$$\tfrac{1}{2}|m| \; ||u-u_k||_2^2 \leq |J(u) - J(u_k) - DJ(u_k)(u-u_k)|$$

$$\leq DJ(u_k)(u-u_k)$$

$$\leq ||DJ(u_k)|| \; ||u-u_k||_2$$

$$||u-u_k||_2 \leq \frac{2}{|m|} ||DJ(u_k)|| \qquad u \varepsilon S_k \; .$$

Since $v \varepsilon S_k$ for all k,

$$||v - u_k||_2 \leq \frac{2}{|m|} ||DJ(u_k)||$$

which was to be shown.

Equation (3.7) does not imply that

$$u_k \to v \text{ in } L_2^m[0,1]$$

because what is known is that

$$DJ(u_k)h_k \to 0$$

which does not guarantee that the operator norm of $DJ(u_k)$ is small. However, since $||DJ(u_k)||$ can be calculated explicitly, (3.7) does provide a useful bound if $||DJ(u_k)||$ is small. For the unconstrained problem, $||DJ(u_k)||$ will be small for large k and (3.7) is a useful estimate.

3.2.1 Linear Quadratic Cost Problems - Equations to be used in the conditional gradient are developed. The second derivative of J(u) is shown to be negative definite which means a bound (3.7) can be obtained. It is explicitly evaluated. The linear quadratic equations are given by (1.9). First the formula for the step size is derived.

Let $$y(t) = \int_0^t e^{A(t-s)} Bh(s)ds \; . \qquad (3.8)$$

Remark 3.7 $\int_0^1 (x_u^T(t)Qy(t)+U^T(t)Rh(t))dt = DJ(u)h.$

Proof $DJ(u)h = \int_0^1 \frac{\partial H}{\partial u}(u)h(t)dt$

$$\frac{\partial H}{\partial u}(u) = \lambda_u^T(t)B + u^T(t)R$$

$$\lambda_u^T(t) = (\int_t^1 e^{-A^T(t-s)}Qx_u(s)ds)^T$$

$$= \int_t^1 x_u^T(s)Qe^{A(s-t)}ds. \quad \text{Thus,}$$

$$\int_0^1 \frac{\partial H}{\partial u}(u)h(t)dt = \int_0^1 \int_t^1 x_u^T(t)Qe^{A(s-t)}dsh(t)dt + \int_0^1 u^T(t)Rh(t)dt \ .$$

Switchthe order of integration of the first integral, exchange the variables s and t and substitute for y(t) using (3.8). It can be seen that,

$$DJ(u)h = \int_0^1 (x_u^T(t)Q\int_0^t e^{A(t-s)}Bh(s)ds+u^T(t)Rh(t))dt$$

$$= \int_0^1 (x_u^T(t)Qy(t)+u^T(t)Rh(t))dt$$

which was to be shown.

Remark 3.8 $D^2J(u)hh = \int_0^1 (y^T(t)Qy(t)+h^T(t)Rh(t))dt.$

Proof: Using the representation of DJ(u)h given by remark 3.7, and recalling the equation for x_{u+h}, the following is obtained.

$$DJ(u+h_2)h_1-DJ(u)h_1 = \int_0^1 (x_{u+h_2}^T(t)Qy_1(t)+(u(t)+h_2(t))^T Rh_1(t) - x_u^T(t)Qy_1(t)+u^T(t)Rh_1(t))dt$$

$$= \int_0^1 (y_2^T(t)Qy_1(t)+h_2^T(t)Rh_1(t))dt$$

where $\quad y_2(t) = \int_0^1 e^{A(t-s)}Bh_2(t)dt.$

Thus, setting $D^2J(u)h_2h_1, = \int_0^1 (y_2^T(t)\ Qy\ (t)+h_2^T(t),Rh,(t))dt,$ it can be seen that $D^2J(u)h_2h_1$ is linear in h_1 for fixed u and h_2 and linear in h_2 for fixed u and h_1. That $D^2J(u)h_2h_1$ is a bounded linear operator for all $u\epsilon\Omega_A$ (in fact for all $u\epsilon L_2^m[0,1]$) follows from the fact that $D^2J(u)h_2h_1$ is independent of u and the definition of the matrices $e^{A(t-s)}$, Q and R where the explicit bounds are derived below. Thus, the remark is proved.

Remark 3.9 $D^2J(u)$ is symmetric, negative definite and bounded below.

Proof:

$$D^2J(u)hh = \int_0^1 (y^T(t)Qy(t) + h^T(t)Rh(t))dt$$

$$\leq \int_0^1 h^T(t)Rh(t)dt$$

$$<0 \qquad \text{for all } h(t) \neq \overline{0}_m \quad \text{a.e.}$$

In particular $D^2J(u)hh \leq r_m||h||_2^2<0$

where r_m is the largest eigenvalue of R which is negative since R is negative definite. Moreover,

$$D^2J(u)hh \geq q_1||y||_2^2 + r_1||h||_2^2$$

where q_1 and r_1 are the smallest eigenvalues of Q and R respectively.

$$||y||_2^2 = \int_0^1 [\int_0^t e^{A(t-s)}Bh(s)ds]^T . [\int_0^t e^{A(t-s)}Bh(s)ds]dt$$

$$\leq \int_0^1 \int_0^t |e^{A(t-s)}Bh(s)|^T_{ds} \int_0^t |e^{A(t-s)}Bh(s)|ds$$

$$= ||\int_0^t |e^{A(t-s)}Bh(s)|ds||_2^2$$

$$\leq ||e^A||_s^2 \cdot ||B||_s^2||h||_2^2$$

$$\leq e^{||A||_s^2} \cdot ||B||_s^2 ||h||_2^2 \ .$$

Therefore,

$$(q_1 e^{||A||_s^2}||B||_s^2+r_1)||h||_2^2 \leq D^2J(u)hh \leq r_m||h||_2^2<0 \text{ for all } h(t)\neq\overline{0}_m \quad \text{a.e.}$$

and the remark is proved.

Note that $D^2J(u)$ is independent of u. Thus, the step size can be calculated as follows

$$g(\alpha) = J(u+\alpha h)$$

$$= J(u) + \alpha DJ(u)h + \tfrac{1}{2}\alpha^2 D^2 J(u)hh$$

and

$$\bar{\alpha} = -\frac{DJ(u)h}{D^2J(u)hh} \tag{3.9}$$

is well defined and positive where u and h come from the conditional gradient sequence. Here $h(t) \neq \bar{0}_m$ a.e. since the conditional gradient sequence would have terminated if $h(t) = \bar{0}_m$ a.e. The step size is

$$\alpha = \min(\bar{\alpha}, 1) .$$

Remark 3.10 The cost functional of the linear quadratic problem (1.9) is strictly concave.

Proof: $x_{\alpha u_1+\beta u_2}(t) = \alpha x_{u_1}(t) + \beta x_{u_2}(t) \quad 0 \leq \alpha \leq 1 , \beta = 1 - \alpha$.

Thus, $J(\alpha u_1+\beta u_2) - (\alpha J(u_1)+\beta J(u_2)) =$

$$-\tfrac{1}{2}\alpha\beta \int_0^1 ((x_{u_1}(t)-x_{u_2}(t))^T Q(x_{u_1}(t)-x_{u_2}(t)) +$$

$$(u_1(t)-u_2(t))R(u_1(t)-u_2(t)))dt>0$$

and the remark is proved.

The results of this section can be used in conjunction with interval analytic techniques developed in chapter 4 to obtain approximate solutions with guaranteed error bounds.

3.2.2 Nonlinear Problems - The nonlinear optimal control problem is difficult to solve via the conditional gradient method not only because it is difficult to solve the associated nonlinear differential equations, but a formula for the step size α_k is not available in general. Moreover, it must be shown that a particular nonlinear problem belongs to class R. If a bound for $||v-u_k||$ is to be obtained, the second derivative must be shown to be negative definite.

Assume the nonlinear problem is in class R. The algorithm can be carried out numerically if a way can be found to obtain α_k so that $J(u_{k+1})>J(u_k)$. The method outlined below is a standard method.

Let $0<\bar{\alpha}_1<\ldots<\bar{\alpha}_N = 1$ be a partition of [0,1]. Then $\alpha_k = \min_i \bar{\alpha}_i$ such that $J(u_k+\bar{\alpha}_i h_k)>J(u_k)$.

Guaranteed error bounds to approximate solutions of nonlinear optimal control problems are derived by techniques developed in chapter 4.

3.2 Example

Let $u\varepsilon\Omega_A'$

$$x'(t) = u(t) \qquad A = 0 \;,\; b = 1$$

$$x(0) = 1$$

$$J(u) = -\tfrac{1}{2}\int_0^1 (x^2(t)+u^2(t))dt \qquad Q = -1 \;,\; R = -1 \quad .$$

Two iterations of the conjugate gradient are carried out. Recall from example 1 of chapter 1 that

$$v(t) = \frac{1}{e^2+1}(e^t-e^2e^{-t})$$

$$x(t) = \frac{1}{e^2+1}(e^t+e^2e^{-t})$$

$$J(v) = -\frac{e^2-1}{2(e^2\_k(} \approx -.\underline{380797078}$$

$$x(t) = 1 + \int_0^t u(s)ds$$

$$H(x,u,\lambda;t) = u(t)\lambda(t) - \tfrac{1}{2}x^2(t) - \tfrac{1}{2}u^2(t)$$

$$\frac{\partial H(u)}{\partial u} = \lambda(t) - u(t)$$

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

$$\lambda(t) = - \int_t^1 x(s)ds$$

From equation (3.9)

$$y(t) = \int_0^t h(s)ds$$

$$\bar{\alpha} = - \frac{\int_0^1 (x(t)y(t)+u(t)h(t))dt}{\int_0^1 (y^2(t)+h^2(t))dt}$$

ITERATION 1 Let $u_0(t) = 0$

$$x_1(t) = 1$$

$$J(u_1) = - \tfrac{1}{2}$$

$$\lambda_1(t) = t - 1$$

$$\frac{\partial H}{\partial u}1(t) = t - 1 \leq 0$$

$$h_1(t) = - 1$$

$$DJ(u_1)h_1 = \int_0^1 (t-1)(-1)dt = \tfrac{1}{2}>0$$

$$y_1(t) = - t$$

$$\alpha_1 = - \frac{\int_0^1 1(-t)dt}{\int_0^1 (t^2+1)dt} = 3/8$$

ITERATION 2 $u_2(t) = u_1(t) + \alpha_1 h_1(t) = - 3/8$

$$x_2(t) = 1 - 3/8t$$

$$J(u_2) = - 13/32 > - \tfrac{1}{2} = J(u_1)$$

$$\lambda_2(t) = -3/16t^2 + t - 13/16$$

$$\frac{\partial H_2}{\partial u^2}(t) = -3/16t^2 - 7/16.$$ Let $a = \frac{8-\sqrt{43}}{3} \approx .48$

$$h_2(t) = \begin{cases} -5/8 & 0 \le t \le a \\ 11/8 & a < t \le 1 \end{cases}$$

$$DJ(u_2)h_2 = 1/8a^3 - a^2 + 7/8a \approx .2 < \tfrac{1}{2} = DJ(u_1)h_1$$

$$y_2(t) = \begin{cases} -5/8t & 0 \le t \le a \\ 11/8t - 2a & a < t \le 1 \end{cases}$$

$$\alpha_2 = \frac{13/16a - a^2 + 1/8a^3}{121/48 - 91/24a + 4a^2 - 43/12a^3} \approx .141$$

4. TWO NUMERICAL METHODS FOR SOLVING OPTIMAL CONTROL PROBLEMS WITH AUTOMATICALLY CALCULATED ERROR BOUNDS USING INTERVAL ANALYSIS

Integral equations and the conditional gradient algorithm have been used to analyze class Q and class R optimal control problems respectively. The numerical methods which will be applied to obtain approximate solutions to class Q and class R problems with guaranteed error bounds is the subject of this chapter.

4.1 Interval Analysis and Optimal Control Problems

The computer is used to obtain approximate solutions with guaranteed error bounds representing the accumulation of discretization error and computer round-off for unconstrained linear quadratic problems and class R optimal control problems. The accumulation of error is accomplished using interval arithmetic computer programs RIA (rounded interval arithmetic), developed for the Control Data Cyber 70 - Model 73 digital computer by Winter (17). He also developed a linear integral equation solver FRED, which yields approximate solutions with guaranteed error bounds. The documentation for RIA and FRED is found in Winter (17). Though Winter (17) develops a nonlinear integral equation solver URYSOHN, its application to unconstrained optimal control problem whose integral equation representation is nonlinear is not possible since the kernel (2.15) is discontinuous. Suggestions for obtaining approximate solutions to (2.15) with guaranteed error bounds is the subject of section 4.2.2.

The computer program FRED is used to obtain a rigorous evaluation of the unconstrained linear quadratic problem. Problems which are solved using the conditional gradient method will require a two stage process the second of which uses the computer programs RIA to obtain a rigorous evaluation of the control problem.

4.2 Unconstrained Problems

The use of FRED yields a rigorous evaluation of the state function. That is, equation (2.8) is solved using FRED. Equation (2.8) is a linear integral equation of the state function x(t). Its rigorous evaluation yields an interval extension function $x^I(t^I)$ which can be used to obtain a rigorous evaluation of $\lambda(t)$ using a first order method (since the first derivative of x(t) is bounded) in conjunction with equation (2.2). The interval extension function $\lambda^I(t^I)$ can be substituted into equation (2.1) to obtain a rigorous evaluation of v(t). Lastly, if the rigorous evaluation of the cost functional is desired,

$$J^I(v^I) = \tfrac{1}{2} \int_0^1 (x^{I^T}(t^I)Q^I x^I(t^I) + v^{I^T}(t^I)R^I v^I(t^I))dt$$

is calculated using the Riemann sum method where the Riemann sum method is defined below.

$$\int_a^b f(t)dt \in \sum_{j=1}^{m} f^I(P_{mj}^I)W_{mj}^I \qquad (4.1)$$

For the above, RIA is used to evaluate the sum and partitions of $[a_j,b_j]$ of [a,b] have the property that

$$[a_j,b_j] = P_{mj} \subset P_{mj}^I .$$

and the weights

$$W_{mj} \subset W_{mj}^I .$$

The solution to the unconstrained nonlinear optimal control problem as transformed into the nonlinear equation (2.15) is

$$y(t) = (x^T(t), \lambda(t))^T$$

which assumes the problem belongs to class Q and that an explicit function $q(x,\lambda;t)$ exists. The functions x and λ are substituted into equation (2.9) and v(t) is obtained. If the cost functional is to be evaluated, then x and v are substituted into

$$J(v) = \int_0^1 L(x,v;t)dt .$$

4.2.1 Linear Quadratic Problems - The computer program FRED can applied to the linear integral equation (2.8) in conjunction with the midpoint rule since the kernel (2.7) is continuous and its first derivative with respect to the integrating factor is bounded. The following two functions and their partials are needed for the implementation of FRED.

$$g(r,t) = k(r,t)e^{Ar}x_0 = \begin{cases} k_1(r,t)e^{Ar}x_0 & 0 \le r \le t \le 1 \\ k_2(r,t)e^{Ar}x_0 & 0 \le t < r \le 1. \end{cases} \qquad (4.3)$$

$$h(r,t,u) = k(r,t)k(u,r) = \begin{cases} k_1(r,t)k_1(u,r) & 0 \le u \le r, t \le 1 \\ k_1(r,t)k_2(u,r) & 0 \le r < u \le t \le 1 \\ k_2(r,t)k_1(u,r) & 0 \le t < u \le r \le 1 \\ k_2(r,t)k_2(u,r) & 0 \le r,t < u \le 1 . \end{cases} \qquad (4.4)$$

The partials needed are $\frac{\partial g}{\partial r}(r,t)$ and $\frac{\partial h}{\partial r}(r,t,u)$ and can be obtained in a straight forward manner. The specific equations used for one dimensional test problems are derived in appendix A.

4.2.2 Nonlinear Problems - The nonlinear unconstrained optimal control problem transformed into a nonlinear integral (2.15) has an inherent discontinuity in its kernel at $s = t$. Moreover, for all nonlinear integral equation representations of nonlinear unconstrained optimal control problems given by Falb and deJong (6), it can be shown that the kernel is discontinuous at $s = t$. If a nonlinear integral equation representation could be found such that the kernel were continuous in s and t, then the method of Winter (17) might be applied to obtain approximate solutions with guaranteed error bounds.

The first suggestion for obtaining a rigorously evaluated $y(t)$ that may work is the following. Consider (2.10) and (2.11). Instead of transforming the optimal control problem into a nonlinear integral equation (2.15), regard it as the two point boundary problem below.

$$\begin{aligned} &y(t) = (x^T(t),\lambda(t))^T \varepsilon R^{2n} \\ &y'(t) = F(y;t) \\ &y_i(0) = x_i(0) \qquad i = 1,\ldots,n \\ &y_i(1) = \lambda_{i-n}(1) \qquad i = n+1,\ldots,2n \ . \end{aligned} \tag{4.5}$$

Using ordinary arithmetic and standard numerical techniques for obtaining approximate solutions to boundary valued problems such as the "shooting" method, approximate the solution to (4.5). Next consider the integral equation (2.15) as the following operator equation.

$$P(y)(t) = y(t) - \int_0^1 k(s,t)F(y;s)ds - c \ . \tag{4.6}$$

It can be shown that,

$$DP(y)h(t) = h(t) - \int_0^1 k(s,t)\frac{\partial F}{\partial y}(y;s)h(s)ds$$

if $\frac{\partial F}{\partial y}$ exists and is continuous in y for fixed s and bounded in [0,1] for fixed y. Moreover, if $D^2P(y)hh$ can be shown to exist or if $\frac{\partial F}{\partial y}(y;s)$ can be shown to satisfy a Lipschitz condition in y for $s \varepsilon [0,1]$, then the

Kantorovich theorem guaranteeing the existence and uniqueness of solutions to (4.6) can be applied to the approximate solution of (4.5). First the Kantorovich theorem is stated without proof, and then its application in conjunction with the approximate solution of (4.5) is outlined.

Let X be a Banach space, B(x,r) be an open ball of radius r centered at x and $\overline{B(x,r)}$ denote its closure. Define the functions w and w^+ by

$$w(h) = \begin{cases} 1 - \sqrt{1-2h}/h & 0<h\leq\frac{1}{2} \\ 1 & h = 0 \end{cases}$$

and

$$w^+(h) = (1+\sqrt{1-2h})/h \qquad 0<h\leq\tfrac{1}{2}$$

Theorem 4.1 (Kantorovich) Assume P is a twice differentiable operator on a Banach space X, $x_0 \varepsilon X$, r>0 and

(1) $(DP(x_0))^{-1}$ exists in the space of bounded linear operators from X to X with $||(DP(x_0))^{-1}||\leq\alpha$,

(2) $||P(x_0)||\leq\beta$,

(3) $||D^2P(x)||\leq \gamma$ for all $x\varepsilon B(x_0,r)$,

(4) $h = \alpha^2\beta\gamma\leq\frac{1}{2}$

(5) $r\geq r_0 = \alpha\beta w(h)$

Then there exists a unique $x^* \varepsilon \overline{B(x_0,r_0)}$ such that $P(x^*) = 0$. If $h<\frac{1}{2}$, then x^* is the unique zero of P in $B(x_0,r)\cap B(x_0,r_1)$ where $r_1 = \alpha\beta w^+(h)$.

It is possible to weaken the differentiability assumption on P and replace condition (3) with the condition that $||DP(x) - DP(y)||\leq ||x-y||$

for all x,y in $B(x_0 r)$.

Let $\bar{y}$ be the approximate solution from (4.5) and use it in theorem 4.1 as x_0. Assume α, β and γ have been rigorously computed and $h = \alpha^2 \beta\gamma \leq \frac{1}{2}$. Then by theorem 4.1, there exists a unique solution y to (4.6) such that $y \varepsilon B(\bar{y}, r_0)$. The rigorously evaluated extension function of the unique solution y becomes

$$y^I(t^I) = \bar{y}^I(t^I) + [-r_0, r_0]. \tag{4.7}$$

Here $\bar{y}(t)$ is exact as a guess to the solution to (4.6) and $\bar{y}^I(t^I)$ need only account for computer roundoff. If $h > \frac{1}{2}$ then a better approximation to the solution of (4.5) is sought.

A second suggestion is to smooth the kernel of (2.15) in the following way. Define,

$$\bar{k}(s,t) = \begin{cases} k_1(s,t) - I_{2n} & 0 \leq s \leq t \leq 1 \\ k_2(s,t) & 0 \leq t < s \leq 1 \end{cases}.$$

Then, (2.15) becomes

$$y(t) = c + \int_0^1 \bar{k}(s,t) F(y;s) ds + \int_0^1 G_0(s,t) F(y;s) \tag{4.8}$$

where $$G_0(s,t) = \begin{cases} I_{2n} & 0 \leq s \leq t \leq 1 \\ 0 & 0 \leq t < s \leq 1 \end{cases}.$$

The first integral is continuous. If a special quadrature rule could be developed for the second integral whose order of convergence is equal to that of the first integral, then an approximate solution to (2.15) might be obtained. Jacoby (10) has done the above for linear integral equations. However, it may be impossible to obtain an approximate solution with guaranteed error bounds; that is, to find a bound for

$$||y_k(t) - y(t)||_n \ .$$

Moreover, a bound for

$$|J(u_k) - J(v)|$$

would have to be obtained if a rigorously evaluated approximate solution to the cost functional is to be found where

$$u_k(t) = q(x_k,\lambda_k;t)$$

and the approximate solution $y_k(t)$ from (4.11) is used in (2.9).

Lastly, consider the following scheme. Suppose $u_k \varepsilon \Omega_A^2$ or $u_k \varepsilon \Omega_n$ is known

(1) Solve for x_k in equations (1.1) forward in time using u_k

(2) Solve for λ_k , where
$\lambda_k'(t) = - \lambda_k(t)\frac{\partial f}{\partial x}(x_k,u_k;t)$ backward
in time since $\lambda(1)$ is given.

(3) $u_{k+1}(t) = q(x_k,\lambda_k;t)$

(4) Replace k by k+1 and begin at step (1).

Consider steps (1) through (4) as a map

$$B:R^m \to R^m \ .$$

If $B(u)$ can be shown to be a contractive map, then approximate solutions with guaranteed error bounds can be obtained.

4.3 Interval Analysis Applied to the Constrained Optimal Control Problem

The conditional gradient method and its application in obtaining rigorously evaluated solutions is outlined next. Whereas the unconstrained optimal control problem was solved assuming a normal condition, this condition was not imposed on the constrained problems though the concavity of the cost functional was assumed. Unconstrained problems whose cost functionals are concave can be rigorously evaluated by the method outlined below.

Rigorously evaluated solutions to class R problems involve a two stage process, a pre-interval analytic stage, and an interval analytic stage. In the pre-interval analytic stage, stage A, an initial guess $u_1 \varepsilon \Omega_A$ is obtained; for example, $u_1(t) \equiv \overline{0}_m$. The conditional gradient method is employed on $J(u)$ using standard arithmetic and differential techniques. If $DJ(u_k)h_k$ is nonpositive, then enter the interval analytic stage, stage B, with $v(t) = u_k(t)$. Otherwise, after it appears that there is little change in the values of succeeding cost functinals, enter stage B if the problem cannot be shown to have a negative definite second derivative $D^2J(u)$. When $D^2J(u)$ is known to be negative definite evaluate

$$\frac{2}{|m|}||DJ(u_k)||$$

If this value is sufficiently small, go to stage B. Otherwise, an improvement to the above is sought by re-entering stage A with u_k, the last control function obtained. If the value

$$\frac{2}{|m|}||DJ(u_k)||$$

shows no significant decrease after a predetermined number of loops back into stage A, go to stage B and a rigorous evaluation of $v(t)$ using (3.7) should be avoided.

Stage B is carried out as follows. The last iterate u_k is assumed to be exact and its interval extension function accounts for round-off error so that $u_k^I(t)$ is obtained using RIA. A rigorous evaluation of $x_k^I(t)$ is obtained using the Riemann sum method (4.1) when the problem is linear quadratic. Otherwise, use an interval version of Euler's method.

The rigorous evaluation of $\lambda_k(t)$ is made next using an interval midpoint rule if the problem is linear quadratic. Otherwise, an interval Euler method is employed. RIA is used to evaluate $\frac{\partial H_k}{\partial u}$ and h_k rigorously. At the same time, if $\frac{2}{|m|}||DJ(u_k)||$ is sufficiently small, evaluate this value using RIA. The formula for the latter value in the linear quadratic case is derived in appendix B.

The value of the interval containing J(v) can be determined by applying the Riemann sum method to the defining integral of $J(u_k)$ and thus

$$J(v) \in J^I(u_k^I) + D^I J(u_k^I)h_k^I \qquad (4.9)$$

using RIA for this calculation. Stage B is stopped unless $\frac{2}{|m|}||J(u_k)||$ is sufficiently small. If it is, then let the interval containing it be denoted as follows,

$$\frac{2}{|m|}||DJ(u_k)|| \subset [d,e] \text{ 2nd } E^I = [-e,e].$$

This means from equation (3.8) 2nd appendix B, that

$$v(t) \subset u_k^I(t) + E^I$$

where the addition is performed using RIA. The entire method is outlined below.

<u>Stage A</u>: Pre-interval analytic

Suppose $u_k \varepsilon \Omega_A'$ or $u_k \varepsilon \Omega_n$ is known.

1. $x_k(t)$ is obtained by solving

$$x_k'(t) = f(x_k, u_k; t)$$

$$x_k(0) = x_o$$

forward in time.

2. $\lambda_k(t)$ is obtained by solving

$$\lambda_k'(t) = - \frac{\partial H}{\partial x}(x_k, u_k, \lambda k; t)$$

$$\lambda_k(1) = \lambda(1)$$

backward in time.

3. Calculate $\frac{\partial H_k}{\partial u}(t) = \lambda_k(t)\frac{\partial f}{\partial u}(x_k, u_k; t)$

4. Calculate the optimal direction,

$$h_k(t) = \text{sgn}\ \frac{\partial H_k}{\partial u}(t) - u_k(t)$$

5. Calculate

$$DJ(u_k)h_k = \int_0^1 \frac{\partial H_k}{\partial u}(t)h_k(t)dt$$

If nonpositive enter stage B with $v(t) = u_k(t)$. Otherwise proceed to step 6.

6. The stepsize α_k is calculated using equation (3.9) if solving a linear quadratic problem or by the method described in section 3.2.2 using (3.10).

7. Set $u_{k+1}(t) = u_k(t) + \alpha_k h_k(t)$

8. The convergence criteria is checked,

$$0 \leq J(u_{k+1}) - J(u_k) \leq \varepsilon_1$$

where ε_1 is a small number. However, since J(u) can be costly to evaluate at each iteration, the succeeding values of control functions is often used as a convergence check. Since u_k is not guaranteed to converge to v, caution should be exercised in using the difference of succeeding control functions as a convergence test.

If the convergence test is met go to stage B unless the problem is known to have a negative definite second derivative $D^2J(u)$ in which case go to step 9. If the convergence test is not met, go to step 1 with u_{k+1} replacing u_k.

9. Check to see if

$$\frac{2}{|m|}||DJ(u_k)|| \leq \varepsilon_2$$

If the check is met, to to stage B. If the value has improved from its last evaluation, make ε_1 smaller and go to step 1 with u_{k+1} replacing u_k. Otherwise, go to stage B. If there is little or no improvement in succeeding values after a predetermined number of loops back to step 1, go to stage B.

Stage B: Interval Analytic Stage

1. Obtain the interval extension function $u_k^I(t)$ using RIA to round up and down to compensate for the finite representation of a number by the computer.

2. Evaluate $x_k^I(t)$ forward in time using equation (4.1) if the problem is linear or an interval Euler method using RIA and error bounds.

3. If stage B was entered as a result of $DJ(u_k)h_k \leq 0$, go to step 8. Otherwise, evaluate $\lambda_k^I(t)$ backward in time using an interval midpoint rule for linear quadratic problems or an interval Euler method by applying RIA and error bounds.

4. Calculate $\frac{\partial H_k^I}{\partial u}(t)$ using RIA.

5. Calculate $h_k^I(t)$ using RIA.

6. Calculate $DJ^I(u_k^I)h_k^I(t)$ by multiplying $\frac{\partial H^I}{\partial u^k}(t)$ and $h_k^I(t)$ using RIA and integrating using (4.1).

7. If the problem is not negative definite or if the $\frac{2}{|m|}||DJ(u_k)||$ was found to be large, go to step 8. Otherwise, calculate

$$\frac{2}{|m|}||\frac{\partial H^I}{\partial u^k}(x_k^I,u_k^I,\lambda_k^I;t)||_2 \qquad (4.10)$$

since $||DJ(u_k)|| \leq ||\frac{\partial H_k}{\partial u}||_2$ (see Appendix B) Let [d,e] denote the interval containing (4.10). Using RIA, perform the following steps

(a) $v^I(t) = u_k^I(t) + [d,e]$.

(b) Calculate $x^I(t)$ using $v^I(t)$ as in step B2.

(c) If $\lambda^I(t)$ is desired, use $x^I(t)$, $v^I(t)$, as in step B3.

(d) Evaluate $J^I(v^I) = \int_0^1 L^I(x^I, v^I; t)dt$ using (4.1) if a rigorous evaluation is necessary.

(e) Stop the process.

8. Calculate $J^I(v^I)$ using (4.9) where $J^I(u_k^I) = \int_0^1 L^I(x_k^I, u_k^I; t)dt$ is evaluated using (4.1).

9. Stop.

Some of the equations and bounds necessary to implement stage A and stage B for linear quadratic problems are found in Appendix B.

BIBLIOGRAPHY

1. Bellman, R., Introduction to Matrix Analysis, McGraw Hill Co., New York, 1970.

2. Bryson, A.E., Denham, W.F., "A Steepest Ascent Method for Solving Optimum Programming Problems", Journal of Applied Mechanics (Trans. ASME, Ser. E) 29, June, 1962.

3. Coddington, E.A., Levinson, N., Theory of Ordinary Differential Equations, McGraw Hill Co., New York, 1970.

4. Demyanov, V., Rubinov, A., Approximate Methods in Optimization Problems, American Elsevier Publishing Co., New York, 1970.

5. Dieudonné, J., Foundations of Modern Analysis, Academic Press, New York, 1969.

6. Falb, P.L., deJong, J.L., Some Successive Approximation Methods in Control and Oscillation Theory. Academic Press, New York, 1969.

7. Halkin, H., "Mathematical Foundations of System Optimization" in Optimization Techniques (G. Leitmann ed.) chapter 6, Academic Press, New York, 1967.

8. Halkin, H., "On the Necessary Conditions for Optimal Control on Nonlinear Systems", Journal D'Analyse Mathématique, 12, 1964.

9. Hasdorff, L., Gradient Optimization and Nonlinear Control, John Wiley and Sons, Inc., New York, 1976.

10. Jacoby, J., Numerical Integration of Linear Integral Equations With Weakly Discontinuous Kernels, Technical Report No. 37, Department of Mathematics, Oregon State University, June 1968.

BIBLIOGRAPHY (Cont'd.)

11. Kelly, H., "Methods of Gradients" in Optimization Techniques With Applications to Aerospace Systems (ed. G. Leitmann), Academic Press, New York, 1962.

12. Krabs, W., Optimal Control and Approximation Theory, (mimeographed seminar notes), Department of Mathematics, Oregon State University, Winter, 1977.

13. Moore, R.E., Interval Analysis, Prentice-Hall, Inc., New Jersey, 1966.

14. Mufti, H.I., Computational Methods in Optimal Control Problems, Springer Verlag, New York, 1970.

15. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F., The Mathematical Theory of Optimal Processes, John Wiley and Sons, Inc., New York, 1962.

16. Rudin, W., Real and Complex Analysis, McGraw Hill, New York, 1974.

17. Winter, L.T., Computer Techniques Yielding Automatic and Rigorous Solutions to Linear and Nonlinear Integral Equations, Technical Report No. 50, Department of Mathematics, Oregon State University, November, 1976.

APPENDICES

APPENDIX A

This appendix contains the equations and bounds needed by Winters (17) one-dimensional linear integral equation solver FRED, as applied to unconstrained linear quadratic optimal control problems. For the one dimensional case

$$x(t) = e^{at}x_o + b \int_0^t e^{a(t-s)}u(s)ds$$

$$\lambda(t) = q\int_0^t e^{-a(t-s)}x(s)ds$$

$$J(u) = \tfrac{1}{2}\int_0^1 [qx^2(t) + pu(t)]dt$$

$q \leq o$ $p < o$ where p is used in place of r so that it is not confused with the integrating factor r.

From (2.5)

$$k_1(r,t) = -\frac{b^2q}{p} e^{at}\cdot e^{ar}\cdot\int_0^r e^{-2as}ds$$

$$= \frac{b^2q}{2ap} e^{at}(e^{-ar}-e^{ar}) \qquad a\neq 0 \qquad (A.1)$$

$$k_1(r,t) = \frac{b^2q}{p} r \qquad a=0 \qquad (A.2)$$

From (2.6)

$$k_2(r,t) = \frac{b^2q}{p} e^{ar}\cdot e^{at}\cdot\int_0^t e^{-2as}ds$$

$$= \frac{b^2q}{2ap} e^{ar}(e^{-at}-e^{at}) \qquad a\neq 0 \qquad (A.3)$$

$$k_2(r,t) = -\frac{b^2q\cdot t}{p} \qquad a=0 \qquad (A.4)$$

$$x(t) - \int_0^1 k(r,t)x(r)dr = e^{at}x_o$$

$$k(r,t) = \begin{cases} k_1(r,t) & 0\leq r\leq t\leq 1 \\ k_2(r,t) & 0\leq t<r\leq 1 \end{cases} \qquad (A.5)$$

Recall $g(r,t) = k(r,t)e^{Ar}x_o$. From (4.5), a≠0 (A.1), and (A.2)

$$g(r,t) = \frac{b^2qx_o e^{at}}{2ap}\begin{cases} e^{at}(e^{-ar}-e^{ar}) & 0\leq r\leq t\leq 1 \\ e^{ar}(e^{-at}-e^{at}) & 0\leq t<r\leq 1 \end{cases} \tag{A.6}$$

If a=0

$$g(r,t) = -\frac{b^2qx_o}{p}\begin{cases} r & 0\leq r\leq t\leq 1 \\ t & 0\leq t<r\leq 1 \end{cases} \tag{A.7}$$

Recall $h(r,t,u) = k(r,t)k(u,r)$. From (4.6), a≠0

$$h(r,t,u) = \frac{b^4q^2}{4a^2p^2}\begin{cases} e^{at}(1-e^{2ar})(e^{-au}-e^{au}) \\ e^{at}(e^{-ar}-e^{ar})^2e^{au} \\ e^{2ar}(e^{-at}-e^{at})(e^{-au}-e^{au}) \\ (1-e^{2ar})(e^{-at}-e^{at})e^{au} \end{cases} \tag{A.8}$$

If a=0

$$h(r,t,u) = \frac{b^4q^2}{p^2}\begin{cases} ru \\ r2 \\ tu \\ tr \end{cases} \tag{A.9}$$

The computer program FRED requires the calculation of

$$G(t) = \int_0^1 g(r,t)dr \quad \text{and}$$

$$H(t,u) = \int_0^1 h(r,t,u)dr$$

and Lipschitz constants L_t and $L_{t,u}$ are needed. Let a≠0, a>0. The Lipschitz constant L_t for $g(r,t)$ is:

If a = o $$L_t = \frac{b^2 q |x_o| (e^{3a} + e^a)}{2p} \qquad \text{(A.10)}$$

$$L_t = \frac{b^2 q |xd}{p} \qquad \text{(A.11)}$$

$$L_{t,u} \leq \frac{b^4 q^2 e^{3a}(e^a+e^{-a})}{2ap^2} \quad \text{for } a\neq 0 \qquad (A.12)$$

since

$$\max \left|\frac{\partial h(r,t,u)}{\partial r}\right| = \max_{0\leq r\leq 1} \frac{b^4 q^2}{4a^2 p^2} \begin{cases} 2ae^{at}e^{2ar}(e^{au}-e^{-au}) \\ 2ae^{at}(e^{ar}+e^{-ar})e^{au} \\ 2ae^{2ar}(e^{at}-e^{-at})(e^{au}-e^{-au}) \\ 2ae^{2ar}(e^{at}-e^{-at})e^{au} \end{cases}$$

$$\leq \frac{b^4 q^2}{4a^2 p^2} \begin{cases} 2ae^{2a}\, e^{a}(e^{a}-e^{-a}) \\ 2a\; e^{a}(e^{a}+e^{-a})e^{a} \\ 2ae^{2a}(e^{a}-e^{-a})(e^{a}-e^{-a}) \\ 2ae^{2a}(e^{a}-e^{-a})e^{a} \end{cases}$$

$$\leq \frac{b^4 q^2 e^{3a}(e^a+e^{-a})}{2ap^2}$$

$$L_{t,u} \leq \frac{2b^4 q^2}{p^2} \quad \text{for } a=0 \qquad (A.13)$$

APPENDIX B

This appendix contains the equations and bounds needed in solving a constrained linear quadratic optimal control as outlined in section 4.3. The formulas not derived in (1.9) which are needed are as follows. Recall for linear quadratic problems λ is a column vector.

$$\frac{\partial H^T}{\partial u}(x,u,\lambda;t) = B^T\lambda(t) + Ru(t) \tag{B.1}$$

$$h(t) = \operatorname{sgn}(B^T\lambda(t) + Ru(t)) - u(t) \tag{B.2}$$

$$DJ(u)h = \int_0^1(|B^T\lambda(t) + Ru(t)| - u^T(t)B^T\lambda(t) - u^T(t)Ru(t))dt \tag{B.3}$$

$$\begin{aligned} y(t) &= e^{At}\int_0^t e^{-As}Bh(s)ds \\ &= e^{At}x_0 + \int_0^t e^{A(t-s)}B\operatorname{sgn}\frac{\partial H}{\partial u}(s)ds - x(u;t) \\ &= \hat{y}(t) - x(u;t) \end{aligned} \tag{B.4}$$

$$\bar{\alpha} = -\frac{\int_0^1(x^T(t)Qy(t) + u^T(t)Rh(t))dt}{\int_0^1(y^T(t)Qy(t) + h^T(t)Rh(t))dt} \tag{B.5}$$

Let $\hat{y}(t) \equiv e^{At}x_0 + \int_0^t e^{A(t-s)}B\operatorname{sgn}\frac{\partial H}{\partial u}(s)ds$. Then

$$x^T(t)Qy(t) = x^T(t)Q\hat{y}(t) - x^T(t)Qx(t) \tag{B.6}$$

$$u^T(t)Rh(t) = u^T(t)Q\operatorname{sgn}\frac{\partial H}{\partial u}(t) - u^T(t)Ru(t) \tag{B.7}$$

$$y^T(t)Qy(t) = \hat{y}(t)Q\hat{y}(t) - 2\hat{y}^T(t)Qx(t) + x^T(t)Qx(t) \tag{B.8}$$

$$h^T(t)Rh(t) = \operatorname{sgn}^T\frac{\partial H}{\partial u}(t)R\operatorname{sgn}\frac{\partial H}{\partial u}(t) - 2\operatorname{sgn}^T\frac{\partial H}{\partial u}(t)Ru(t) + u^T(t)Ru(t) \tag{B.9}$$

Suppose $||\cdot||$ is any matrix norm, then

$$||e^{At}|| \leq e^{||At||}$$

$$||e^{At} - I - \sum_{j=1}^{k} (At)^j|| = ||\sum_{j=k+1}^{\infty} (At)^j||$$

$$\leq \frac{||At||}{(k+1)!}[1 + \frac{||At||}{k+2} + \frac{||At||^2}{(k+2)(k+3)} + \ldots]$$

$$\leq \frac{||At||}{(k+1)!} e^{||At||}$$

Therefore, an interval extension function for e^{At} can be found since

$$e^{At} \varepsilon\ I^I + \sum_{j=1}^{k} \frac{(A^I t^I)^j}{j!} + E^I \tag{B.10}$$

where $E^I = [-\frac{||At||}{(k+1)!}^{k+1} e^{||At||}, \frac{||At||}{(k+1)!}^{k+1} e^{||At||}]$

Equation (3.6) is used in evaluation of (3.4) where

$$DJ(u_k)h_k = \int_0^1 \frac{\partial H_k}{\partial u}(t)h_k(t)dt$$

is obtained from theorem 1.1. Thus, from (B.3)

$$0 \leq J(v) - J(u_k) \leq \sum_{i=1}^{m} \int_0^1 |\frac{\partial H_k}{\partial u_i}(t)|dt - \int_0^1 \frac{\partial H_k}{\partial u}(t)u_k(t)dt$$

$$\leq \sum_{i=1}^{m} \int_0^1 |\lambda_k^T(t)B^i + u_k^T(t)R^i| - \int_0^1 (\lambda_k^T(t)Bu_k(t) + u_k^T(t)Ru_k(t))dt \tag{B.11}$$

Let $h \varepsilon \Omega_A$. The operator norm $||DJ(u_k)||$ is calculated as follows

$$\sup_{||h||_2 \leq 1} |DJ(u_k)h| \leq \sup_{||h||_2 \leq 1} \int_0^1 |\frac{\partial H_k}{\partial u}(t)||h(t)|dt$$

$$\leq \sup_{||h||_2 \leq 1} ||\frac{\partial H_k}{\partial u}(t)||_2 ||h||_2$$

$$\leq \quad ||\lambda_k^T B||_2 + ||u_k^T R||_2 \qquad \text{(B.12)}$$

Letting r_1 denote the largest eigenvalue of R

$$||v-u_k||_2 \leq \frac{2}{|r_1|} ||DJ(u_k)||$$

$$\leq \frac{2}{|r_1|} ||\frac{\partial H_k}{\partial u}||_2 \quad \text{or}$$

$$\leq \frac{2}{|r_1|} (||\lambda_k^T B||_2 + ||u_k^T R||_2) \qquad \text{(B.13)}$$

The first order midpoint rule that can be used to solve for $\lambda(t)$ and is given as follows:

$$|\int_a^b f(t)dt - f((a+b)/2)\cdot(b-a)| \leq ((b-a)^2/4)\cdot L$$

where L is the Lipschitz constant for f on [a,b]. The $\sup_{a\leq s\leq b} f'(s)$ could replace L in the above. Let c([a,b]) denote the center of the interval [a,b]. Using the notation of (4.1)

$$\int_a^b f(t)dt \in \sum_{j=1}^{m} f(c(P_{mj}^I))w(P_{mj}^I) + \sum_{j=1}^{m} ((W_{mj}^I)^2/4) f^{I\,'}(P_{mj}^I) \qquad \text{(B.14)}$$

where the first sum is performed using ordinary arithmetic and the second sum is performed using RIA. To add the two quantities, the first sum is considered exact and rounded up and down to account for the finite bit representation of a number on the computer and added using RIA to the interval containing the second sum.

Recall that from (2.2)

$$\lambda(t) = e^{-A^T t} \int_t^1 e^{A^T s} Qx(s)ds$$

Let $f(t) = e^{A^T t}Qx(t)$. Then

$$f'(t) = A^T e^{A^T t} Qx(t) + e^{A^T t} Q(Ax(t) + Bu(t))$$

Therefore,

$$f^{I'}(P_{mj}^I) = A^{T^I} e^{A^{T^I} P_{mj}^I} Q^I x^I(P_{mj}^I) + e^{A^{T^I} P_{mj}^I} Q^I (A^I x^I(P_{mj}^I) + B^I u^I(P_{mj}^I) \tag{B.15}$$