WELL, Klaus Hinrich, 1940-
MODIFIED QUASILINEARIZATION METHOD FOR
OPTIMAL CONTROL PROBLEMS WITH BOUNDED
STATE VARIABLES.

Rice University, Ph.D., 1972
Engineering, aeronautical

University Microfilms, A XEROX Company, Ann Arbor, Michigan
INFORMATION TO USERS

This dissertation was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
A Xerox Education Company
RICE UNIVERSITY

MODIFIED QUASILINEARIZATION METHOD
FOR OPTIMAL CONTROL PROBLEMS
WITH BOUNDED STATE VARIABLES

by

KLAUS H. WELL

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Thesis Director's Signature

Houston, Texas
March, 1972
PLEASE NOTE:

Some pages may have

indistinct print.

Filmed as received.

University Microfilms, A Xerox Education Company
ABSTRACT

Modified Quasilinearization Method

for Optimal Control Problems with Bounded State Variables

by

KLAUS H. WELL

This thesis considers the numerical solution of optimal control problems involving a cost function $J$ subject to differential constraints, a state variable inequality constraint, and terminal constraints. The problem is to find the state $x(t)$, the control $u(t)$, and the parameter $\pi$ in such a way that the cost functional is minimized, while the constraints are satisfied to a predetermined accuracy.

In contrast to penalty function techniques and transformation techniques, the state variable inequality constraint is handled in a direct manner. That is, for the time interval for which the trajectory of the system lies on the state boundary, the control vector is determined in such a way that the state boundary is satisfied to a predetermined accuracy.

To facilitate the numerical solution on digital computers, the actual time $\theta$ is replaced by the normalized time $t$, defined in such a way that each of the subarcs composing the extremal arc has a normalized time length $\Delta t = 1$. In this way, variable-time corner conditions and variable-time terminal conditions are transformed into fixed-time corner conditions and fixed-time terminal conditions. The actual times $\theta_1$, $\theta_2$, $\tau$ at which (i) the state boundary is entered, (ii) the state boundary is exited,
and (iii) the terminal boundary is reached are regarded to be components of the parameter \( \pi \) being optimized.

A modified quasilinearization algorithm is developed. Its main property is the descent property in the performance index \( R \), the cumulative error in the constraints and the optimum conditions. Modified quasilinearization differs from ordinary quasilinearization because of the inclusion of the scaling factor (or stepsize) \( \alpha \) in the system of variations. The stepsize is determined by a one-dimensional search on the performance index \( R \).

Since the first variation \( \delta R \) is negative, the decrease in \( R \) is guaranteed if \( \alpha \) is sufficiently small. Convergence to the solution is achieved when \( R \) becomes smaller than some preselected value.

In order to start the algorithm, some nominal functions \( x(t), u(t), \pi \) and nominal multipliers \( \lambda(t), \rho(t), \sigma, \mu \) must be chosen. In a real problem, the selection of the nominal functions can be made on the basis of physical considerations. Concerning the nominal multipliers, no useful guidelines have been available thus far. In this paper, an auxiliary minimization algorithm for selecting the multipliers optimally is presented: the performance index \( R \) is minimized with respect to \( \lambda(t), \rho(t), \sigma, \mu \). Since the functional \( R \) is quadratically dependent on the multipliers, the resulting variational problem is governed by optimality conditions which are linear and, therefore, can be solved without difficulty.

Two numerical examples demonstrate the feasibility as well as the rapidity of convergence of the technique developed in this thesis.
ACKNOWLEDGEMENT

The author is indebted to Professor A. Miele for suggesting the topic and helpful discussions. Furthermore, he is thankful to Dr. R.R. Iyer for checking the accuracy of the numerical results. He would like also to thank his wife, Diane for her patience and understanding during the time this thesis progressed. Last but certainly not least, he would like to thank Mrs. Sandi Peppeard for the excellent typing of the manuscript.

This research was supported by the Office of Scientific Research, Office of Aerospace Research, United States Air Force, Grant No. AF-AFOSR-72-2185.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Statement of the Problem</td>
<td>5</td>
</tr>
<tr>
<td>3. First-Order Conditions</td>
<td>11</td>
</tr>
<tr>
<td>4. Approximate Methods</td>
<td>13</td>
</tr>
<tr>
<td>5. Modified Quasilinearization Algorithm</td>
<td>15</td>
</tr>
<tr>
<td>6. Auxiliary Minimization Algorithm</td>
<td>27</td>
</tr>
<tr>
<td>7. Numerical Examples</td>
<td>33</td>
</tr>
<tr>
<td>8. Remark</td>
<td>50</td>
</tr>
<tr>
<td>9. Discussion and Conclusions</td>
<td>52</td>
</tr>
<tr>
<td>References</td>
<td>56</td>
</tr>
</tbody>
</table>
1. Introduction

Over the past ten years, much effort has been spent in the study of optimal control problems involving state variable inequality constraints. In general, two approaches have been used to solve these problems. The first approach is indirect in nature, in that additional state variables are introduced and/or the cost functional is modified suitably to account for the state variable inequality constraints. The second approach is direct in nature in that, whenever any portion of the trajectory of the system lies on the state boundary, the control vector is determined in such a way that the state boundary is satisfied.

In the area of indirect approaches, the most widely used approach is a penalty function scheme: the original, constrained optimization problem is transformed into an equivalent, unconstrained optimization problem through the introduction of additional state variables and/or suitable modification of the cost functional. In this connection, penalty function techniques were developed in Refs. 1-2 in conjunction with the steepest descent algorithm and in Ref. 3 in conjunction with the generalized Newton-Raphson algorithm. The advantage of this approach is that the number and sequence of subarcs composing the extremal arc need not be known a priori. The disadvantage is that the rapidity of convergence to the solution of the original, constrained optimization problem depends heavily on a skillfull

---

1 Within the context of this thesis, the term unconstrained optimization is employed to denote a problem without state variable inequality constraints, while the term constrained optimization is employed to denote a problem with the state variable inequality constraints.
choice of the penalty constants; this drawback is fundamental, since no
clear-cut criterion exits for choosing penalty constants.

Another indirect approach, developed in Ref. 4, uses a Valentine-type
representation of the state variable inequality constraint. In this technique,
the dimension of the state vector is increased, a new controller is introduced, and
the optimization takes place in this new and larger state space with respect to
the new controller. The advantage of this approach is that the number
and sequence of subarcs composing the extremal arc need not be known
a priori. The disadvantage is that, if $k$ is the order to the state variable
inequality constraint\(^2\), the dimension of the state vector is increased by $k$.

In the area of direct approaches, some of the earlier work was done
in Refs. 5-7, and more recent work was done in Refs. 8-10. The common
element of Refs. 5-10 is that the number and sequence of subarcs composing
the extremal arc must be known a priori. On the state variable boundary,
the inequality constraint is employed with the equality sign. The left-hand
side of the resulting equality constraint is differentiated as many times as
needed ($k$ times) until the control variable appears explicitly. Then, the
equality constraint and its first $k-1$ derivatives are employed as corner
conditions for the state variable boundary, while the last derivative is
used as a control variable constraint to be satisfied everywhere on the
state boundary.

The present thesis belongs to the area of direct approaches and extends to
constrained optimization problems the modified quasilinearization algorithm

\(^2\) A state variable inequality constraint is defined to be of order $k$ if the
$k$th time derivative of the left-hand side of the constraint relation is the
first to contain the control variable explicitly.
developed in Ref. 11 for unconstrained optimization problems. This algorithm incorporates two important features: (a) a clear-cut criterion for choosing the scaling factor (or stepsize) $\alpha$ of the system of variation and (b) a clear-cut criterion for determining the times at which the state boundary is entered, the state boundary is exited, and the terminal boundary is reached.

Feature (a) is achieved by means of a one-dimensional search on the performance index $R$, the cumulative error in the constraints and the optimum conditions. Since the first variation $\delta R$ is negative, the decrease in $R$ is guaranteed if $\alpha$ is sufficiently small.

Feature (b) is achieved by replacing the actual time $\theta$ with the normalized time $t$, defined in such a way that each of the subarcs composing the extremal arc has a normalized time length $\Delta t = 1$. In this way, variable-time corner conditions and variable-time terminal conditions are transformed into fixed-time corner conditions and fixed-time terminal conditions. The actual times at which (i) the state boundary is entered, (ii) the state boundary is exited, and (iii) the terminal boundary is reached are regarded to be components of the parameter $\tau$ being optimized (Ref. 12).

In order to start the algorithm, a judicious choice of the nominal functions $x(t), u(t), \tau$ and the nominal multipliers $\lambda(t), \rho(t), \sigma, \mu$ is necessary. In a real problem, the selection of the nominal functions can be made on the basis of physical considerations. Concerning the nominal multipliers, the method developed in Ref. 11 for unconstrained optimization problems is extended here to constrained optimization problems. The initial multipliers $\lambda(t), \rho(t), \sigma, \mu$ are determined in such a way that the performance index $R$ is
minimized. Since R is quadratically dependent on the multipliers, the resulting variational problem is governed by optimality conditions which are linear and, therefore, can be solved without difficulty.
2. Statement of the Problem

The purpose of this thesis is to study the minimization of the cost functional

\[ I = \int_0^\tau f_*(x, u, \theta) d\theta + [g_*(x, \theta)]_\tau \]

with respect to the state \( x(\theta) \), the control \( u(\theta) \), and the parameter \( \tau \) which satisfy the differential constraint

\[ \frac{dx}{d\theta} - \varphi_*(x, u, \theta) = 0 \]  

the state variable inequality constraint\(^3\)

\[ L_*(x, \theta) \geq 0 \]

and the boundary conditions

\[ \langle x \rangle_0 = \text{given} \quad [\psi_*(x, \theta)]_\tau = 0 \]

In the above equations, the functions \( f_* \), \( g_* \), \( L_* \) are scalar, the function \( \varphi_* \) is an \( n \)-vector, and the function \( \psi_* \) is a \( q \)-vector. The independent variable is the actual time \( \theta \) (a scalar), and the dependent variables are the state \( x \) (an \( n \)-vector), the control \( u \) (an \( m \)-vector), and the parameter \( \tau \) (a scalar).

At the initial time \( \theta = 0 \), \( n \) scalar relations are specified. At the final time \( \theta = \tau \), \( q \) scalar relations are specified, where \( q \leq n \) if \( \tau \) is fixed and \( q \leq n+1 \) if \( \tau \) is free.

2.1. Composition of the Extremal Arc. The state variable inequality constraint (3) plays a fundamental role in the problem under consideration.\(^3\)

\(^3\) The function \( L_*(x, \theta) \) is assumed to be linear in \( x \) and \( \theta \) (see Section 8).
This being the case, we divide the extremal arc into subarcs where (3) is satisfied with the inequality sign almost everywhere and subarcs where (3) is satisfied with the equality sign. Assuming that the equality sign in (3) is not compatible with (4-1) at the initial point and (4-2) at the final point, we postulate that the extremal arc includes three subarcs, specifically,

\[ L_*(x, \theta) \geq 0 \quad , \quad 0 \leq \theta \leq \theta_1 \]
\[ L_*(x, \theta) = 0 \quad , \quad \theta_1 \leq \theta \leq \theta_2 \]
\[ L_*(x, \theta) \geq 0 \quad , \quad \theta_2 \leq \theta \leq \tau \]

In Eqs. (5), \( \theta_1 \) denotes the time at which the central subarc (5-2) is entered, and \( \theta_2 \) denotes the time at which the central subarc (5-2) is exited. For the initial subarc (5-1), the inequality sign holds everywhere, except at \( \theta = \theta_1 \), where equality holds\(^4\). Analogously, for the final subarc (5-3), the inequality sign holds everywhere, except at \( \theta = \theta_2 \), where equality holds\(^4\).

2.2. State Variable Boundary. Here, we consider the subarc on the state variable boundary

\[ L_*(x, \theta) = 0 \]

and assume it to be of order \( k \). We denote by \( L_*(x, \theta) \), \( \dot{L}_*(x, \theta) \), \( \ddot{L}_*(x, \theta) \), ..., \( \overset{k-1}{L}_*(x, \theta) \), \( \overset{k}{L}_*(x, \theta) \) the successive total derivatives of \( L_*(x, \theta) \), that is,

\[ \overset{1}{L}_*(x, \theta) = \frac{dL_*(x, \theta)}{d\theta} \]
\[ \overset{2}{L}_*(x, \theta) = \frac{d\overset{1}{L}_*(x, \theta)}{d\theta} \]

\[ \overset{k-1}{L}_*(x, \theta) = \frac{d\overset{k-2}{L}_*(x, \theta)}{d\theta} \]
\[ \overset{k}{L}_*(x, \theta) = \frac{d\overset{k-1}{L}_*(x, \theta)}{d\theta} \]

\( ^4 \)This is due to the fact that the state variable \( x \) is continuous.
By definition, the control $u$ does not appear explicitly in the first $k-1$ derivatives, while it appears explicitly in the $k$th derivative.

Next, we define the $k$-vector function\footnote{Within the context of this paper, all vectors are column vectors. The symbol $^T$ denotes matrix transposition.}

$$M_*(x, \theta) = [L_*(x, \theta), L_*^1(x, \theta), \ldots, L_*^{k-1}(x, \theta)]^T$$ \hfill (8)

and the scalar function

$$N_*(x, u, \theta) = L_*^k(x, u, \theta)$$ \hfill (9)

With this understanding, we see that the subarc on the state variable boundary \footnote{The function $M_*(x, \theta)$ is assumed to be linear in $x$ and $\theta$ (see Section 8).} can be represented through the equality constraints

$$M_*(x, \theta) = 0 \quad , \quad N_*(x, u, \theta) = 0$$ \hfill (10)

We note that the components of the vector equation (10-1) are successive first integrals of the scalar equation (10-2). Hence, (10-1) need not be applied everywhere but only at the initial point or at the final point of the subarc on the state variable boundary (6). This being the case, (10-1) can be replaced with

$$[M_*(x, \theta)]_{\theta_1} = 0 \quad \text{or} \quad [M_*(x, \theta)]_{\theta_2} = 0$$ \hfill (11)

2.3. Time Normalization. To facilitate the development of the necessary conditions for optimality as well as to facilitate the implementation of the modified quasilinearization algorithm on a digital computer, we replace the actual time $\theta$ with the normalized time $t$. The latter is defined in such
a way that the normalized time length of each of the subarcs composing the extremal arc is \( \Delta t = 1 \). Therefore, in normalized form, \( t = 1 \) denotes the time at which the state boundary (6) is entered, \( t = 2 \) denotes the time at which the state boundary (6) is exited, and \( t = 3 \) denotes the time at which the terminal boundary (4-2) is reached.

The following linear relations allow one to pass from the normalized time \( t \) to the actual time \( \theta \):

\[
\begin{align*}
\theta &= \theta_1 t , & 0 \leq t \leq 1 \\
\theta &= \theta_1 + (\theta_2 - \theta_1)(t - 1) , & 1 \leq t \leq 2 \\
\theta &= \theta_2 + (\tau - \theta_2)(t - 2) , & 2 \leq t \leq 3 
\end{align*}
\]

(12)

and imply that

\[
\begin{align*}
\frac{d\theta}{dt} &= \theta_1 , & 0 \leq t \leq 1 \\
\frac{d\theta}{dt} &= \theta_2 - \theta_1 , & 1 \leq t \leq 2 \\
\frac{d\theta}{dt} &= \tau - \theta_2 , & 2 \leq t \leq 3 
\end{align*}
\]

(13)

The fact that the normalized corner times are fixed (\( t=1 \) and \( t=2 \)) and the normalized final time is fixed (\( t=3 \)) does not cause any loss of generality in the problem. The actual corner times (\( \theta=\theta_1 \) and \( \theta=\theta_2 \)) and the actual final time (\( \theta=\tau \)) can be regarded to be components of a vector parameter \( \pi \) to be optimized. Specifically, \( \pi \) is given by

\[
\pi = [\theta_1, \theta_2]^T \quad \text{or} \quad \pi = [\theta_1, \theta_2, \tau]^T
\]

(14)

where (14-1) holds if \( \tau \) is fixed and (14-2) holds if \( \tau \) is free.
2.4. Definitions. In addition to the normalized time \( t \) and the parameter \( \pi \), we define the following functions:

\[
\begin{align*}
\varphi(x, u, \pi, t) &= \left\{ \begin{array}{ll}
\theta_2 f_*(x, u, \theta), & 0 \leq t \leq 1 \\
\left( \frac{t}{2} - \theta_1 \right) f_*(x, u, \theta), & 1 \leq t \leq 2 \\
\left( \tau - \theta_2 \right) f_*(x, u, \theta), & 2 \leq t \leq 3
\end{array} \right. \\
\varphi(x, u, \pi, t) &= \left\{ \begin{array}{ll}
\theta_2 \varphi_*(x, u, \theta), & 0 \leq t \leq 1 \\
\left( \frac{t}{2} - \theta_1 \right) \varphi_*(x, u, \theta), & 1 \leq t \leq 2 \\
\left( \tau - \theta_2 \right) \varphi_*(x, u, \theta), & 2 \leq t \leq 3
\end{array} \right. 
\end{align*}
\]

and

\[
\begin{align*}
L(x, \pi, t) &= L_*(x, \theta), & 1 \leq t \leq 2 \\
M(x, \pi, t) &= M_*(x, \theta), & 1 \leq t \leq 2 \\
N(x, u, \pi, t) &= N_*(x, u, \theta), & 1 \leq t \leq 2
\end{align*}
\]

and

\[
g(x, \pi, t) = \tilde{g}_*(x, \theta), \quad \psi(x, \pi, t) = \psi_*(x, \theta)
\]

2.5. New Formulation. In the light of the previous discussion, problem (1)-(4) can be reformulated as follows: Minimize the cost functional

\[
I = \int_{0}^{3} f(x, u, \pi, t)dt + [g(x, \pi)]_3
\]

with respect to the state \( x(t) \), the control \( u(t) \), and the parameter \( \pi \) which satisfy the differential constraint

\footnote{The dot denotes derivative with respect to the normalized time \( t \).}
\[ \dot{x} - \varphi(x, u, \pi, t) = 0 \quad , \quad 0 \leq t \leq 3 \] (20)

the control variable equality constraint

\[ N(x, u, \pi, t) = 0 \quad , \quad 1 \leq t \leq 2 \] (21)

and the multipoint conditions

\[ (x)_0 = \text{given} \quad , \quad [M(x, \tau)]_1 = 0 \quad , \quad [\psi(x, \tau)]_3 = 0 \] (22)

In the above equations, the functions \( f, g, N \) are scalar, the function \( \varphi \) is an \( n \)-vector, the function \( M \) is a \( k \)-vector, and the function \( \psi \) is a \( q \)-vector.

The independent variable is the normalized time \( t \) (a scalar), the dependent variables are the state \( x \) (an \( n \)-vector), the control \( u \) (an \( m \)-vector), and the parameter \( \pi \) (a \( p \)-vector), where \( \pi \) is given by (14-1) if \( \tau \) is fixed \((p=2)\) and \( \pi \) is given by (14-2) if \( \tau \) is free \((p=3)\).

At the initial time \( t = 0 \), the \( n \) scalar relations (22-1) are specified. At the intermediate time \( t = 1 \), the \( k \) scalar relation (22-2) are specified, where \( k \) is the order of the state variable boundary. At the final time \( t = 3 \), the \( q \) scalar relations (22-3) are specified, where \( q \leq n \) if \( \tau \) is given and \( q \leq n+1 \) if \( \tau \) is free. Of course, implicit in the previous formulation is the assumption that

\[ L(x, \pi, t) \geq 0 \quad , \quad 0 \leq t \leq 1 \text{ and } 2 \leq t \leq 3 \] (23)

\[ L(x, \pi, t) = 0 \quad , \quad 1 \leq t \leq 2 \]
3. First-Order Conditions

From calculus of variations, it is known that the previous problem is one of the Bolza type. It can be recast as that of minimizing the augmented cost functional

\[ J = \int_{0}^{3} (\lambda^T \dot{x} + H) dt + (F)_{1} + (G)_{3} \]  

(24)

subject to (20)-(22), where

\[ F = \sigma^T M, \quad G = g + \mu^T \psi, \quad H = f - \lambda^T \phi + \rho N \]  

(25)

and

\[ \rho = 0, \quad 0 \leq t \leq 1 \text{ and } 2 \leq t \leq 3 \]

\[ \rho \neq 0, \quad 1 < t < 2 \]  

(26)

In the above expressions, \( \lambda(t) \) is a variable Lagrange multiplier (an n-vector), \( \rho(t) \) is a variable Lagrange multiplier (a scalar), \( \sigma \) is a constant Lagrange multiplier (a k-vector), and \( \mu \) is a constant Lagrange multiplier (a q-vector).

Upon integrating by parts the term \( \lambda^T \dot{x} \), we see that the augmented cost functional (24) can be written as

\[ J = \int_{0}^{3} (-\lambda^T \dot{x} + H) dt - (\lambda^T x)_{0} - [(\lambda^T x)_{1} - [(\lambda^{+} + \lambda^{-})^T x - F]_{1} - [(\lambda^T x)_{2} + (\lambda^T x + G)_{3} \]  

(27)

If standard techniques of the calculus of variations or optimal control theory are employed, we see that the solution extremizing (27) must satisfy the optimality conditions

\[ \text{The subscript minus denotes conditions preceding a corner point, and the subscript plus denotes conditions following a corner point.} \]
\[ \dot{\lambda} - H_x = 0 \quad 0 \leq t \leq 3 \quad (28) \]

and

\[ H_u = 0 \quad , \quad 0 \leq t \leq 3 \quad (29) \]

and

\[ \int_0^3 H \pi \, dt + (F_1 + (G_1) \pi_3 = 0 \quad (30) \]

and

\[ (\lambda_+ - \lambda_- - F_x)_1 = 0 \quad , \quad (\lambda_+ - \lambda_-)_2 = 0 \quad , \quad (\lambda + G_x)_3 = 0 \quad (31) \]

Summarizing, we seek functions \( x(t), u(t), \pi \) and multipliers \( \lambda(t), \rho(t), \sigma, \mu \) which satisfy the constraints (20)-(22) and the optimality conditions (28)-(31).
4. **Approximate Methods**

In general, the differential system (20)-(22) and (28)-(31) is nonlinear, Consequently, approximate methods must be used to seek sequentially a solution. In this connection, define the norm of the vector \( \nu \) to be

\[
\text{norm}(\nu) = \nu^T \nu
\]  

(32)

Then, the performance indexes

\[
P = \int_0^3 \text{norm}(\dot{x} - \phi) dt + \int_1^2 \text{norm}(N) dt + \text{norm}(M)_1 + \text{norm}(\psi)_3
\]

(33)

and

\[
Q = \int_0^3 \text{norm}(\dot{\lambda} - H_x \dot{x}) dt + \int_0^3 \text{norm}(H_u) dt + \text{norm} \left[ \int_0^3 H \ dt + (F \pi)_1 + (G \pi)_3 \right] \\
+ \text{norm}(\lambda_+ - \lambda_-) \_1 + \text{norm}(\lambda_+ - \lambda_-) \_2 + \text{norm}(\lambda + G_{\pi}) \_3
\]

(34)

measure the errors in the constraints and the optimum conditions, respectively. For the exact optimal solution, one must have

\[
P = 0 \quad , \quad Q = 0
\]

(35)

For an approximation to the optimal solution, one must have

\[
P \leq \epsilon_1 \quad , \quad Q \leq \epsilon_2
\]

(36)

where \( \epsilon_1 \) and \( \epsilon_2 \) are small, preselected numbers. Alternatively, (36) can be replaced by

\[
R \leq \epsilon_3
\]

(37)
where the performance index

\[ R = P + Q \]  \hfill (38)

denotes the cumulative error in the constraints and the optimum conditions.

In (37), \( \epsilon_3 \) is a small, preselected number. Note that satisfaction of Ineq. (37) implies satisfaction of Ineqs. (36) if one chooses

\[ \epsilon_1 = \epsilon_2 = \epsilon_3 \]  \hfill (39)
5. **Modified Quasilinearization Algorithm**

Here, we present a modification of the quasilinearization algorithm which has a descent property in the performance indexes $P$, $Q$, $R$. Consider the nominal functions and the multipliers $x(t)$, $u(t)$, $\pi$, $\lambda(t)$, $\rho(t)$, $\sigma$, $\mu$ and the varied functions and multipliers $\tilde{x}(t)$, $\tilde{u}(t)$, $\tilde{\pi}$, $\tilde{\lambda}(t)$, $\tilde{\rho}(t)$, $\tilde{\sigma}$, $\tilde{\mu}$ such that

$$
\begin{align*}
\tilde{x}(t) &= x(t) + \Delta x(t) \quad \tilde{u}(t) = u(t) + \Delta u(t) \quad \tilde{\pi} = \pi + \Delta \pi \\
\tilde{\lambda}(t) &= \lambda(t) + \Delta \lambda(t) \quad \tilde{\rho}(t) = \rho(t) + \Delta \rho(t) \quad \tilde{\sigma} = \sigma + \Delta \sigma \quad \tilde{\mu} = \mu + \Delta \mu
\end{align*}
$$

(40) (41)

Here, $\Delta x(t)$, $\Delta u(t)$, $\Delta \pi$, $\Delta \lambda(t)$, $\Delta \rho(t)$, $\Delta \sigma$, $\Delta \mu$ denote the perturbations of $x(t)$, $u(t)$, $\pi$, $\lambda(t)$, $\rho(t)$, $\sigma$, $\mu$ about the nominal values.

5.1. **First Variation.** The passage from the nominal functions to the varied functions causes the performance indexes $P$, $Q$, $R$, to change. To first order, we see that

$$
\delta R = \delta P + \delta Q
$$

(42)

where

$$
\begin{align*}
\delta P &= 2\int_{0}^{3} (\tilde{x} - \varphi)^{T} \delta(\tilde{x} - \varphi) dt + 2\int_{1}^{2} N\delta N dt + 2(M)^{T} \delta(M)_{1} + 2(\psi)^{T} \delta(\psi)_{3}
\end{align*}
$$

(43)

and

$$
\begin{align*}
\delta Q &= 2\int_{0}^{3} (\tilde{\lambda} - H_x)^{T} \delta(\tilde{\lambda} - H_x) dt + 2\int_{0}^{3} H_{u}^{T} \delta H_{u} dt + 2\left[ \int_{0}^{3} H_{u} dt + (F_{x})_{1} + (G_{x})_{3} \right]^{T} \delta \int_{0}^{3} H_{u} dt + (F_{x})_{1} + (G_{x})_{3} \\
&\quad + 2(\lambda_{+} - \lambda_{-} - F_{x})^{T} \delta(\lambda_{+} - \lambda_{-} - F_{x})_{1} + 2(\lambda_{+} - \lambda_{-})^{T} \delta(\lambda_{+} - \lambda_{-})_{2}^{2} + 2(\lambda + G_{x})^{T} \delta(\lambda + G_{x})_{3}
\end{align*}
$$

(44)
5.2. **Special Variations.** Next, consider the system of variations defined by the linearized equations

\[ \delta(x - \phi) + \alpha(x - \phi) = 0 \quad , \quad 0 \leq t \leq 3 \]
\[ \delta N + \alpha N = 0 \quad , \quad 1 \leq t \leq 2 \]  

(45)

and

\[ \delta(x)_0 = 0 \quad , \quad \delta(M)_1 + \alpha(M)_1 = 0 \quad , \quad \delta(\psi)_3 + \alpha(\psi)_3 = 0 \]  

(46)

and the linearized optimality conditions

\[ \delta(\dot{\lambda} - H_x) + \alpha(\dot{\lambda} - H_x) = 0 \quad , \quad 0 \leq t \leq 3 \]  

(47)

and

\[ \delta H_u + \alpha H_u = 0 \quad , \quad 0 \leq t \leq 3 \]  

(48)

and

\[ \delta \left[ \int_0^3 \frac{H}{\pi} dt + (F_{\pi})_1 + (G_{\pi})_3 \right] + \alpha \left[ \int_0^3 \frac{H}{\pi} dt + (F_{\pi})_1 + (G_{\pi})_3 \right] = 0 \]  

(49)

and

\[ \delta(\lambda_+ - \lambda_- - F_{x})_1 + \alpha(\lambda_+ - \lambda_- - F_{x})_1 = 0 \]
\[ \delta(\lambda_+ - \lambda_-)_2 + \alpha(\lambda_+ - \lambda_-)_2 = 0 \]
\[ \delta(\lambda + G_{x})_3 + \alpha(\lambda + G_{x})_3 = 0 \]  

(50)
where $\alpha$ is a scaling factor (or stepsize) in the range

$$0 \leq \alpha \leq 1$$  \hspace{1cm} (51)

The fact that these variations are admissible to first order is proven in Section 5.7.

5.3. **Descent Property.** When the variations defined by (45)-(50) are employed, the first variations of the performance indexes $P$ and $Q$ become

$$\delta P = -2\alpha P \ , \ \delta Q = -2\alpha Q$$  \hspace{1cm} (52)

with the implication that

$$\delta R = -2\alpha R$$  \hspace{1cm} (53)

If the nominal functions and multipliers are nonextremal, the relation $R > 0$ holds, so that $\delta R < 0$. Hence, for $\alpha$ sufficiently small, it is guaranteed that

$$R < R$$  \hspace{1cm} (54)

This is the basic descent property of the algorithm defined by Eqs. (45)-(50).

5.4. **System of Variations.** In explicit form, the linearized equations (45)-(46) can be written as

$$\Delta \dot{x} - \phi_x \Delta x - \phi_u \Delta u - \phi_{\pi} \Delta \pi + \phi(\dot{x} - \phi) = 0 \ , \ \ 0 \leq t \leq 3$$

$$N_x \Delta x + N_u \Delta u + N_{\pi} \Delta \pi + \alpha N = 0 \ , \ \ 1 \leq t \leq 2$$  \hspace{1cm} (55)
and

\[(\Delta x)_0 = 0 , \ (M_x^T \Delta x + M_{\pi}^T \Delta \pi + \alpha M)_1 = 0 , \ (\psi_x^T \Delta x + \psi_{\pi}^T \Delta \pi + \alpha \psi)_3 = 0 \quad (56)\]

and the linearized optimality condition (47)-(50) can be written as

\[
\Delta \hat{\lambda} - H_{xx}^T \Delta x - H_{xu}^T \Delta u - H_{x\pi}^T \Delta \pi - H_{x\lambda}^T \Delta \lambda - H_{x\rho}^T \Delta \rho + \alpha (\hat{\lambda} - H_{x} \Delta \lambda) = 0 , \ 0 \leq t \leq 3 \quad (57)
\]

and

\[
H_{ux}^T \Delta x + H_{uu}^T \Delta u + H_{u\pi}^T \Delta \pi + H_{u\lambda}^T \Delta \lambda + H_{u\rho}^T \Delta \rho + \alpha H_{u} \Delta \lambda = 0 , \ 0 \leq t \leq 3 \quad (58)
\]

and

\[
\int_0^3 (H_{xx}^T \Delta x + H_{xu}^T \Delta u + H_{x\pi}^T \Delta \pi + H_{x\lambda}^T \Delta \lambda + H_{x\rho}^T \Delta \rho) dt + (F_{xx}^T \Delta x + F_{xu}^T \Delta u + F_{x\pi}^T \Delta \pi + F_{x\lambda}^T \Delta \lambda + F_{x\rho}^T \Delta \rho)_1 + (G_{xx}^T \Delta x + G_{xu}^T \Delta u + G_{x\pi}^T \Delta \pi + G_{x\lambda}^T \Delta \lambda + G_{x\rho}^T \Delta \rho)_3 = 0 \quad (59)
\]

and

\[
(\Delta \lambda_+ - \Delta \lambda_- - F_{xx}^T \Delta x - F_{x\pi}^T \Delta \pi)_1 + (\lambda_+ - \lambda_- - F_{x} x)_1 = 0
\]

\[
(\Delta \lambda_+ - \Delta \lambda_-)_2 + \alpha (\lambda_+ - \lambda_-)_2 = 0 \quad (60)
\]

\[
(\Delta \lambda + G_{xx}^T \Delta x + G_{x\pi}^T \Delta \pi + G_{x\lambda}^T \Delta \lambda)_3 + \alpha (\lambda + G_{x})_3 = 0
\]

In solving Eqs. (55)-(60), one must remember that

\[
\Delta \rho = 0 , \quad 0 \leq t \leq 1 \text{ and } 2 \leq t \leq 3
\]

\[
\Delta \rho \neq 0 , \quad 1 \leq t \leq 2 \quad (61)
\]
For $\alpha = 1$, Eqs. (55)-(60) become identical with those of ordinary quasilinearization. While modified quasilinearization exhibits the descent property (54), this is not necessarily the case with ordinary quasilinearization. This means that, if Eqs. (55)-(60) are employed with $\alpha = 1$, the performance index $R$ may actually increase when passing from the nominal functions and multipliers to the varied functions and multipliers.

5.5. **Coordinate Transformation.** To simplify the problem, we introduce the auxiliary variables

$$A(t) = \Delta x(t)/\alpha , \quad B(t) = \Delta u(t)/\alpha , \quad C = \Delta \pi/\alpha$$

(62)

$$D(t) = \Delta \lambda(t)/\alpha , \quad E(t) = \Delta \phi(t)/\alpha , \quad S = \Delta \sigma/\alpha , \quad Z = \Delta \mu/\alpha$$

(63)

With these variables, the linearized equations (55)-(56) become

$$\dot{A} - \dot{\varphi} T_{xx} A - \dot{\varphi} T_{uu} B - \varphi T_{\pi \pi} C + \dot{x} - \varphi = 0 , \quad 0 \leq t \leq 3$$

$$N_{xx}^T A + N_{uu}^T B + N_{\pi \pi}^T C + N = 0 , \quad 1 \leq t \leq 2$$

(64)

and

$$\begin{align*}
(A)_{0} &= 0 , \quad (M_{xx}^T A + M_{\pi \pi}^T C + M)_{1} = 0 , \quad (\psi_{xx}^T A + \psi_{\pi}^T C + \psi)_{3} = 0
\end{align*}$$

(65)

and the linearized optimality conditions (57)-(60) become

$$\dot{D} - H_{xx}^T A - H_{uu}^T B - H_{\pi \pi}^T C + H_{\pi \lambda}^T D - H_{\pi \rho}^T E + \dot{\lambda} - H_{x} = 0 , \quad 0 \leq t \leq 3$$

(66)
and
\[ H_{ux}^T A + H_{uu}^T B + H_{u\pi}^T C + H_{u\lambda}^T D + H_{u\rho}^T E + H_u^T H = 0, \quad 0 \leq t \leq 3 \] (67)

and
\[ \int_0^3 (H_{x\pi}^T A + H_{x\mu}^T B + H_{x\pi}^T C + H_{x\lambda}^T D + H_{x\rho}^T D)dt + (F_{x\pi}^T A + F_{x\mu}^T C + F_{x\rho}^T S) \]
\[ + (G_{x\pi}^T A + G_{x\pi}^T C + G_{x\lambda}^T Z) + \int_0^3 H_{\pi}^T dt + (F_{1\pi} + G_{1\pi}^T) = 0 \] (68)

and
\[ (D_+ - D_- - F_{x\pi}^T A - F_{x\pi}^T C - F_{x\rho}^T S + \lambda_+ - \lambda_- - F_{1\pi}) = 0 \]
\[ (D_+ - D_- + \lambda_+ - \lambda_-)_2 = 0 \] (69)
\[ (D + G_{x\pi}^T A + G_{x\pi}^T C + G_{x\lambda}^T Z + \lambda + G_{x}^T) = 0 \]

In solving Eqs. (64)-(69), one must remember that
\[ E = 0, \quad 0 \leq t \leq 1 \text{ and } 2 \leq t \leq 3 \]
\[ E \neq 0, \quad 0 \leq t \leq 2 \] (70)

The differential system (64)-(69) is linear and nonhomogeneous in the functions
A(t), B(t), C, D(t), E(t), S, Z and can be solved without assigning a value to
the stepsizes \( \alpha \). The selection of \( \alpha \) is done \textit{a posteriori} in such a way that the
descent requirement (54) is enforced.

5.6. \textbf{Remark.} Simple manipulations, omitted for the sake of brevity,
show that Eqs. (64-2) and (67) imply that
\[ B = -(H_u^T)^{-1}(W + H_u^T E), \quad 0 \leq t \leq 3 \] (71)

where

\[ E = 0, \quad 0 \leq t \leq 1 \text{ and } 2 \leq t \leq 3 \] (72)

\[ E = \left[N_u^T (H_u^T - H_u^T)^{-1} [U - N_u^T (H_u^T)^{-1} W]\right], \quad 1 \leq t \leq 2 \]

and where

\[ U = N_x^T A + N_\pi^T C + N \]

\[ W = H_u^T A + H_u^T C + H_u^T D + H_u \] (73)

Equations (71)-(72) replace Eqs. (64-2) and (67) in the computational procedure which follows.

5.7. Solution Technique. We integrate the previous differential system \( n + p + k + 1 \) times using a forward integration scheme in combination with the method of particular solutions (Refs. 13-17). In each integration (subscript \( i \)), we assign a different set of values to the \( n \)-vector \( D(0) \), the \( p \)-vector \( C \), and the \( k \)-vector \( S \), specifically,

\[ D_i(0) = [\delta_{i1}, \delta_{i2}, \ldots, \delta_{in}]^T \]

\[ C_i = [\delta_{i1(n+1)}, \delta_{i(n+2)}, \ldots, \delta_{i(n+p)}]^T \] (74)

\[ S_i = [\delta_{i1(n+p+1)}, \delta_{i(n+p+2)}, \ldots, \delta_{i(n+p+k)}]^T \]

where \( i = 1, 2, \ldots, n + p + k + 1 \) and where the Kronecker delta \( \delta_{ij} \) is such that
\[ \delta_{ij} = 1 \text{ if } i = j \]  
\[ \delta_{ij} = 0 \text{ if } i \neq j \]  
(75)

With the above vectors specified, the differential system is integrated forward employing (a) Eqs. (64-1), (65-1), (66), (69-1), (69-2), (71), (72) and bypassing (b) Eqs. (65-2), (65-3), (68), (69-3). In this way, one obtains the functions

\[ A_i(t), B_i(t), C_i, \quad i = 1, 2, \ldots, n + p + k + 1 \]  
(76)

\[ D_i(t), E_i(t), S_i, \quad i = 1, 2, \ldots, n + p + k + 1 \]  
(77)

Next, we introduce the \( n + p + k + 1 \) undetermined, scalar constants \( k_i \) and form the linear combinations

\[ A(t) = \Sigma k_i A_i(t), \quad B(t) = \Sigma k_i B_i(t), \quad C = \Sigma k_i C_i \]  
(78)

\[ D(t) = \Sigma k_i D_i(t), \quad E(t) = \Sigma k_i E_i(t), \quad S = \Sigma k_i S_i \]  
(79)

where the summation is taken over the index \( i \). Then, we inquire whether, by an appropriate choice of the constants \( k_i \) and the components of the parameter \( Z \), these linear combinations can satisfy all the differential equations and boundary conditions. By simple substitution, it can be verified that (78)-(79) satisfy (a) providing

\[ \Sigma k_i = 1 \]  
(80)

and satisfy (b) providing
\[
[\sum k_i (M_{x_1}^T A_1 + M_{\pi_1}^T C_1) + M]_1 = 0
\]
\[
[\sum k_i (\psi_{x_1}^T A_1 + \psi_{\pi_1}^T C_1) + \psi]_3 = 0
\]
\[
[\sum k_i (D_1 + G_{x_1}^T A_1 + G_{x_1}^T C_1) + G_{x_1}^T Z + \lambda + G_{x_3}]_3 = 0
\]

and
\[
\sum_{i=0}^{3} k_i \int_0^3 (H_{\pi_x}^T A_1 + H_{\pi_\mu}^T B_1 + H_{\pi\pi}^T C_1 + H_{\pi\lambda}^T D_1 + H_{\pi\rho}^T E_1) dt
\]
\[
+ \sum_{i=0}^{3} k_i (F_{\pi_x}^T A_1 + F_{\pi_\mu}^T C_1 + F_{\pi\pi}^T S_1) + \sum_{i=0}^{3} k_i (G_{\pi_x}^T A_1 + G_{\pi_\mu}^T C_1)_3
\]
\[
+ (G_{\pi_\mu}^T Z)_3 + \int_0^3 H dt + (F_{\pi_1}) + (G_{\pi_3}) = 0
\]

The linear system (80)-(82) is equivalent to \( n + p + k + 1 + q \) scalar equations, in which the unknowns are \( n + p + k + 1 \) scalar constants \( k_i \) and the \( q \) components of the parameter \( Z \). After the constants \( k_i \) and the components of the parameter \( Z \) are known, the functions \( A(t) \), \( B(t) \), \( C \) are computed with (78) and the functions \( D(t) \), \( E(t) \), \( S \) are computed with (79). In this way, the linear, multipoint boundary-value problem is solved.

5.8. **Determination of the Stepsize.** After combining Eqs. (40)-(41) and (62)-(63), we obtain the relations

\[
\ddot{x}(t) = x(t) + \alpha A(t), \quad \ddot{u}(t) = u(t) + \alpha B(t), \quad \ddot{\pi} = \pi + \alpha C
\]

\[
\ddot{\lambda}(t) = \lambda(t) + \alpha D(t), \quad \ddot{\varphi}(t) = \rho(t) + \alpha E(t), \quad \ddot{\sigma} = \sigma + \alpha S, \quad \ddot{\mu} = \mu + \alpha Z
\]
which constitute a one-parameter family of solutions, the parameter being the stepsize $\alpha$. For this one-parameter family, the performance index (38) becomes a function of the form

$$\tilde{R} = \tilde{R}(\alpha)$$  \hspace{1cm} (85)

At $\alpha = 0$, the slope of this function is negative and is given by

$$\tilde{R}'(0) = -2\tilde{R}(0)$$  \hspace{1cm} (86)

The function (85) exhibits a relative minimum with respect to $\alpha$, that is, a point where

$$\tilde{R}'(\alpha) = 0$$  \hspace{1cm} (87)

This point can be determined by means of a one-dimensional search (for example, using quadratic interpolation, cubic interpolation, or quasilinearization). Ideally, this procedure should be used iteratively until the modulus of the slope satisfies the following inequality:

$$|\tilde{R}'(\alpha)| \leq \varepsilon_4$$  \hspace{1cm} (88)

where $\varepsilon_4$ is a small, preselected number.

Since the rigorous determination of $\alpha$ might take excessive time on a computer, one might renounce solving Eq. (87) with a particular degree of precision and determine the stepsize in a noniterative fashion. To this effect, we first assign the value

$$\alpha = 1$$  \hspace{1cm} (89)
to the stepsize. This is the value which would solve Eq. (87) exactly, should the functional (19) be quadratic and the constraints (20)-(22) be linear. Of course, the stepsize (89) is acceptable only if

\[ \tilde{R}(\alpha) < \tilde{R}(0) \]  

(90)

Otherwise, the previous value of \( \alpha \) must be replaced by some smaller value in the range (51) (for example, using a bisection process) until Eq. (90) is met. This is guaranteed by the descent property (54).

Two supplementary precautions must be taken: (i) the stepsize must be such that

\[ \tilde{L}(\alpha, t) \geq 0 , \quad 0 \leq t \leq 1 \text{ and } 2 \leq t \leq 3 \]  

(91)

where \( \tilde{L}(\alpha, t) \) is the function resulting by combining the left-hand side of the state variable inequality constraint (23-1) and (83); and (ii) the stepsize must be such that

\[ 0 \leq \tilde{\tau}_1(\alpha) \leq \tilde{\tau}_2(\alpha) \leq \tilde{\tau}(\alpha) \]  

(92)

where \( \tilde{\tau}(\alpha) = \tau \) for problems where \( \tau \) is fixed and \( \tilde{\tau}(\alpha) \neq \tau \) for problems where \( \tau \) is free.

In summary, the stepsize must be chosen so that Ineq. (90)-(92) are satisfied. Should any violation occur, then a smaller value of \( \alpha \) must be employed and can be obtained, for example, with a bisection process, starting from \( \alpha = 1 \).

5.9. Summary of the Algorithm. In the light of the previous discussion, we summarize the modified quasilinearization algorithm as follows.
(a) Assume nominal functions $x(t)$, $u(t)$, $\pi$ and multipliers $\lambda(t)$, $\rho(t)$, $\sigma$, $\mu$ consistent with the initial condition (22-1) and the state variable inequality constraint (23-1).

(b) For the nominal functions, compute the coefficients appearing in the linearized equations (64)-(69).

(c) Solve the linear multipoint boundary-value problem (64)-(69) with the method of particular solutions (Section 5.7). Obtain the functions $A(t)$, $B(t)$, $C, D(t)$, $E(t)$, $S$, $Z$.

(d) Using the functions in (c), consider the one-parameter family (83)-(84) and compute the stepsize $\alpha$ by a one-dimensional search on the performance index $\tilde{R}(\alpha)$. To this effect, perform a bisection process on $\alpha$, starting from $\alpha = 1$, until Ineqs. (90)-(92) are satisfied.

(e) Once the stepsize $\alpha$ is known, compute the varied functions $\tilde{x}(t)$, $\tilde{u}(t)$, $\tilde{\pi}$, $\tilde{\lambda}(t)$, $\tilde{\rho}(t)$, $\tilde{\sigma}$, $\tilde{\mu}$ with Eqs. (83)-(84).

(f) With the varied functions known, the iteration is completed. The varied functions $\tilde{x}(t)$, $\tilde{u}(t)$, $\tilde{\pi}$, $\tilde{\lambda}(t)$, $\tilde{\rho}(t)$, $\tilde{\sigma}$, $\tilde{\mu}$ become the nominal functions $x(t)$, $u(t)$, $\pi$, $\lambda(t)$, $\rho(t)$, $\sigma$, $\mu$ for the next iteration. That is, return to (a) and iterate the algorithm.

(e) The algorithm is terminated when the stopping condition (37) is satisfied.
6. **Auxiliary Minimization Algorithm**

In order to start the modified quasilinearization algorithm some
nominal values for the functions $x(t), u(t), \pi, \lambda(t), \rho(t), \sigma, \mu$ must be
chosen. In a real problem, the selection of the state $x(t)$, the control
$u(t)$, and the parameter $\pi$ can be made on the basis of physical considerations.
Regarding the multipliers $\lambda(t), \rho(t), \sigma, \mu$ the following mathematical
argument is proposed.

The cumulative performance index $R$ includes the components $P$ and $Q$
associated with the satisfaction of the constraints and the optimum conditions,
respectively. For given $x(t), u(t), \pi$, the performance index $P$, given by
Eq. (33), is known; on the other hand, the performance index $Q$, given by
Eq. (34), depends quadratically on $\lambda(t), \rho(t), \sigma, \mu$. A logical selection of
the Lagrange multipliers $\lambda(t), \rho(t), \sigma, \mu$ is that which gives to $Q$ (an hence $R$)
the smallest possible value for given $x(t), u(t), \pi$. Hence, we formulate an
auxiliary minimization problem.

6.1. **Auxiliary Minimization Problem.** Let the performance index $Q$,
given by Eq. (34), be rewritten as

$$Q = \int_0^3 \text{norm}(\xi) dt + \int_0^3 \text{norm}(\gamma) dt + \text{norm}(\epsilon) + \text{norm}(\xi) + \text{norm}(\tau) + \text{norm}(\zeta)$$  \hspace{1cm} (93)

where $\xi(t), \gamma(t), \epsilon, \xi, \tau, \zeta$ measure the errors in the optimality conditions
(28)-(31). Therefore, these errors are defined by

$$\dot{\lambda} - f_x + \phi_x \lambda - \rho N_x - \beta = 0, \hspace{1cm} 0 \leq t \leq 3$$  \hspace{1cm} (94)
and

\[ f_u - \varphi_u \lambda + \rho N_u - \gamma = 0, \quad 0 \leq t \leq 3 \quad (95) \]

and

\[ \int_0^3 \left( (f \pi - \varphi \frac{\lambda}{\pi} + \rho N \frac{\lambda}{\pi}) dt + (M \sigma) \right) \left( g \mu + \psi \frac{\mu}{\pi} \right) \pi - \epsilon = 0 \quad (96) \]

and

\[ (\lambda_+ - \lambda_- - M \frac{\sigma}{\pi} - \xi)_{1} = 0, \quad (\lambda_+ - \lambda_- - \eta)_{2} = 0, \quad (\lambda + g \frac{\mu}{\pi} + \psi \frac{\mu}{\pi} - \zeta)_{3} = 0 \quad (97) \]

With these definitions, we formulate the following minimal problem: Find the multipliers \( \lambda(t), \rho(t), \sigma, \mu \) and the functions \( \beta(t), \gamma(t), \epsilon, \xi, \eta, \zeta \) which minimize the function (93) subject to the constraints (94)-(97). In solving the above problem, one must remember that

\[ \rho = 0, \quad 0 \leq t \leq 1 \text{ and } 2 \leq t \leq 3 \quad (98) \]
\[ \rho \neq 0, \quad 1 \leq t \leq 2 \]

6.2. First-Order Conditions. If standard techniques of the calculus of variations or optimal control theory are employed, one must introduce multipliers \( \beta(t), \gamma(t), \epsilon, \xi, \eta, \zeta \) associated with the constraints (94)-(97). The nature of the optimality conditions is such that (i) the following relations hold (Ref. 11):

\[ \beta(t) = \beta(t), \quad \gamma(t) = \gamma(t), \quad 0 \leq t \leq 3 \]
\[ \epsilon = \epsilon, \quad \xi = \xi, \quad \eta = \eta, \quad \zeta = \zeta \quad (99) \]
and (ii) the multiplier $\beta_*(t)$ is continuous at $t = 1$ and $t = 2$. This being the case, the multipliers $\beta_*(t)$, $\gamma_*(t)$, $\varepsilon_*$, $\xi_*$, $\eta_*$, $\zeta_*$ can be readily eliminated, and the optimality conditions take the following form:

$$\dot{\beta} - \varphi_X^T \beta + \varphi_u^T \gamma + \varphi_{\pi}^T \varepsilon = 0, \quad 0 \leq t \leq 3$$ (100)

and

$$N_X^T \beta - N_u^T \gamma - N_{\pi}^T \varepsilon = 0, \quad 1 \leq t \leq 2$$ (101)

with

$$(M_X^T \xi - M_{\pi}^T \varepsilon)_1 = 0, \quad (\psi_X^T \zeta + \psi_{\pi}^T \varepsilon)_3 = 0$$ (102)

and

$$(\beta)_{0} = 0, \quad (\beta - \varepsilon)_1 = 0, \quad (\beta - \eta)_2 = 0, \quad (\beta + \zeta)_3 = 0$$ (103)

The differential system (94)-(103) is linear and nonhomogeneous in the multipliers $\lambda(t)$, $\sigma(t)$, $\mu$ and the functions $\beta(t)$, $\gamma(t)$, $\varepsilon$, $\xi$, $\eta$, $\zeta$.

6.3. **Remark.** Simple manipulations, omitted for the sake of brevity, show that Eqs. (95) and (101) imply that

$$\gamma = f_u - \varphi_u \lambda + \rho N_u, \quad 0 \leq t \leq 3$$ (104)

where

$$\rho = 0, \quad 0 \leq t \leq 1 \text{ and } 2 \leq t \leq 3$$

$$\rho = (N_u^T N_u)^{-1} [N_X^T \beta - N_{\pi}^T \varepsilon - N_u^T (f_u - \varphi_u \lambda)], \quad 1 \leq t \leq 2$$ (105)
Equations (104)-(105) replace Eqs. (95) and (101) in the computational procedure which follows.

6.4. Solution Technique. We integrate the previous differential system \( n + p + k + 1 \) times using a forward integration scheme in combination with the method of particular solutions (Refs. 13-17). In each integration (subscript \( i \)), we assign a different set of values to the \( n \)-vector \( \lambda(0) \), the \( p \)-vector \( \epsilon_i \), and the \( k \)-vector \( \sigma_i \), specifically,

\[
\lambda_i(0) = \begin{bmatrix} \delta_{i1} & \delta_{i2} & \cdots & \delta_{in} \end{bmatrix}^T
\]

\[
\epsilon_i = \begin{bmatrix} \delta_{i(n+1)} & \delta_{i(n+2)} & \cdots & \delta_{i(n+p)} \end{bmatrix}^T
\]

\[
\sigma_i = \begin{bmatrix} \delta_{i(n+p+1)} & \delta_{i(n+p+2)} & \cdots & \delta_{i(n+p+k)} \end{bmatrix}^T
\]  

(106)

where \( i = 1, 2, \ldots, n + p + k + 1 \) and where the Kronecker delta \( \delta_{ij} \) is such that

\[
\delta_{ij} = 1 \quad \text{if} \quad i = j
\]

\[
\delta_{ij} = 0 \quad \text{if} \quad i \neq j
\]  

(107)

With the above vectors specified, the differential system is integrated forward employing (a) Eqs. (94), (97-1), (97-2), (100), (103), (104), (105) and bypassing (b) Eqs. (96), (97-3), (102). In this way, one obtains the functions

\[
\lambda_i(t), \rho_i(t), \sigma_i, \quad i = 1, 2, \ldots, n + p + k + 1
\]

(108)

\[
\beta_i(t), \gamma_i(t), \epsilon_i, \xi_i, \eta_i, \zeta_i, \quad i = 1, 2, \ldots, n + p + k + 1
\]

(109)

Next, we introduce the \( n + p + k + 1 \) undetermined, scalar constants \( k_i \) and form the linear combinations
\[
\lambda(t) = \sum_{i} \lambda_{i}(t), \quad \rho(t) = \sum_{i} \rho_{i}(t), \quad \sigma = \sum_{i} \sigma_{i}
\]  
(110)

\[
\psi(t) = \sum_{i} \psi_{i}(t), \quad \gamma(t) = \sum_{i} \gamma_{i}(t), \quad \epsilon = \sum_{i} \epsilon_{i}
\]  
(111)

\[
\xi = \sum_{i} \xi_{i}, \quad \eta = \sum_{i} \eta_{i}, \quad \zeta = \sum_{i} \zeta_{i}
\]  
(112)

where the summation is taken over the index \( i \). Then, we inquire whether, by an appropriate choice of the constants \( k_{i} \) and the components of the multiplier \( \mu \), these linear combinations can satisfy all the differential equations and boundary conditions. By simple substitution, it can be verified that (110)-(112) satisfy (a) providing

\[
\sum_{i} k_{i} = 1
\]  
(113)

and satisfy (b) providing

\[
[\sum_{i} (M_{x}^{T} \xi_{i} - M_{\pi}^{T} \epsilon_{i})]_{1} = 0
\]

\[
[\sum_{i} (\psi_{x}^{T} \zeta_{i} + \psi_{\pi}^{T} \epsilon_{i})]_{3} = 0
\]  
(114)

\[
[\sum_{i} (\lambda_{i} - \zeta_{i}) + \psi_{x} \mu + g_{x}]_{3} = 0
\]

and

\[
\sum_{i} \int_{0}^{3} (\varphi_{x} \lambda_{i} + \rho_{x} N_{x}) dt + \sum_{i} (M_{x} \sigma_{i} - \epsilon_{i})_{1} + (\psi_{x} \mu)_{3} + \int_{0}^{3} f_{\pi} dt + (g_{x})_{3} = 0
\]  
(115)

The linear system (113)-(115) is equivalent to \( n + p + k + 1 + q \) scalar equations, in which the unknowns are \( n + p + k + 1 \) scalar constants \( k_{i} \) and the \( q \) components...
of the multiplier $\mu$. After the constants $k_i$ and the components of the multiplier $\mu$ are known, the multipliers $\lambda(t), \varrho(t), \sigma$ are computed with (110) and the functions $\beta(t), \gamma(t), \varepsilon, \xi, \eta, \zeta$ are computed with (111)-(112).

In this way, the linear, multipoint boundary-value problem is solved.

6.5. Summary of the Algorithm. In the light of the previous discussion, we summarize the auxiliary minimization algorithm as follows.

(a) Assume nominal functions $x(t), u(t), \pi$ consistent with the initial condition (22-1) and the state variable inequality constraint (23-1).

(b) For the nominal functions, compute the coefficients appearing in the linear equations (94)-(103).

(c) Solve the linear multipoint boundary-value problem (94)-(103) using the method of particular solutions (Section 6.4). Obtain the multipliers $\lambda(t), \varrho(t), \sigma, \mu$ and the functions $\beta(t), \gamma(t), \varepsilon, \xi, \eta, \zeta$. 
7. **Numerical Examples**

In this section, two numerical examples are presented. The first is a fixed-final-time problem, and the second is a free-final-time problem. For simplicity, all symbols employed here denote scalar quantities.

**Convergence** is achieved when a solution consistent with the inequality

\[ R \leq 10^{-8} \tag{116} \]

is found. Conversely, **nonconvergence** is defined by means of the inequalities

\[ N \geq 50 \quad \text{or} \quad N_s \geq 10 \tag{117} \]

Here, \( N \) is the iteration number, and \( N_s \) is the number of bisections of the stepsize \( \alpha \) required to satisfy Ineq. (54); these bisections are started from \( \alpha = 1 \). Satisfaction of Ineq. (117-1) indicates divergence or extreme slowness of convergence; in turn, satisfaction of Ineq. (117-2) indicates extreme smallness of the variations. Both situations are undesirable.

Computations were performed at Rice University on a Burroughs B-5500 digital computer. Double-precision arithmetic was used, and the algorithm was programmed in FORTRAN IV. The interval of integration was divided into 60 steps, that is, 20 steps for each of the three subintervals

\[ 0 \leq t \leq 1 , \quad 1 \leq t \leq 2 , \quad 2 \leq t \leq 3 \tag{118} \]

The differential systems were integrated using Hamming's modified predictor-corrector method with a special Runge-Kutta procedure to start the integration routine (Ref. 18). The definite integrals \( I, P, Q, R \) were computed using Simpson's rule.
7.1. Fixed-Final-Time Example. Consider the problem of minimizing the cost functional

\[ I = \int_0^1 (x^2 + u^2) d\theta \]  \hspace{1cm} (119)

with respect to the state \( x(\theta) \) and the control \( u(\theta) \) which satisfy the differential constraint

\[ \frac{dx}{d\theta} = x^2 - u \quad , \quad 0 \leq \theta \leq 1 \]  \hspace{1cm} (120)

the state variable inequality constraint

\[ x \geq 0.9 \quad , \quad 0 \leq \theta \leq 1 \]  \hspace{1cm} (121)

and the boundary conditions

\[ x(0) = 1 \quad , \quad x(1) = 1 \]  \hspace{1cm} (122)

The solution obtained bypassing Ineq. (122) is given in Ref. 19 and violates Ineq. (122) over the interval \( 0.1 \leq \theta \leq 0.9 \). This being the case, we postulate that the extremal arc includes three subarcs, specifically

\[ x \geq 0.9 \quad , \quad 0 \leq \theta \leq \theta_1 \text{ and } \theta_2 \leq \theta \leq 1 \]  \hspace{1cm} (123)

\[ x = 0.9 \quad , \quad \theta_1 \leq \theta \leq \theta_2 \]

When the normalized time \( t \) is employed, problem (119)-(122) is reformulated as that of minimizing the cost functional
\[ I = \int_{0}^{1} \theta_1 (x^2 + u^2) \, dt + \int_{1}^{2} (\theta_2 - \theta_1) (x^2 + u^2) \, dt + \int_{2}^{3} (1 - \theta_2) (x^2 + u^2) \, dt \] (124)

with respect to the state \( x(t) \), the control \( u(t) \), and the parameter components \( \theta_1, \theta_2 \) which satisfy the differential constraint

\[
\begin{align*}
\dot{x} &= \theta_1 (x^2 - u), & 0 \leq t \leq 1 \\
\dot{x} &= (\theta_2 - \theta_1) (x^2 - u), & 1 \leq t \leq 2 \\
\dot{x} &= (1 - \theta_2) (x^2 - u), & 2 \leq t \leq 3
\end{align*}
\] (125)

the control variable equality constraint

\[ (\theta_2 - \theta_1) (x^2 - u) = 0, \quad 1 \leq t \leq 2 \] (126)

and the multipoint conditions

\[ x(0) = 1, \quad x(1) = 0.9, \quad x(3) = 1 \] (127)

Of course, implicit in the above formulation is the assumption that

\[
\begin{align*}
x &\geq 0.9, & 0 \leq t \leq 1 \text{ and } 2 \leq t \leq 3 \\
x &= 0.9, & 1 \leq t \leq 2
\end{align*}
\] (128)

In this problem,

\[ n = 1, \quad m = 1, \quad p = 2, \quad k = 1, \quad q = 1 \] (129)

Since \( n + p + k + 1 = 5 \), five particular solutions are needed for each iteration.
We assume the nominal state

\[
x(t) = 1 - 0.1t, \quad 0 \leq t \leq 1
\]

\[
x(t) = 0.9, \quad 1 \leq t \leq 2
\]

\[
x(t) = 0.8 + 0.1t, \quad 2 \leq t \leq 3
\]

(130)

the nominal control

\[
u(t) = (12 - 7t + t^2)/6, \quad 0 \leq t \leq 3
\]

(131)

and the nominal parameter components

\[v_1 = 0.1, \quad v_2 = 0.9
\]

(132)

We observe that these nominal functions (labeled \(N = -1\)) are consistent with the multipoint conditions (127) and with the state variable inequality constraint (128-1) over the initial and terminal subarcs. On the other hand, they violate the differential constraint (125) and the control variable equality constraint (126).

**Auxiliary Minimization Algorithm.** Starting with the Nominal functions (130)-(132), we employed the auxiliary minimization algorithm of Section 6 in order to generate the optimal initial multipliers \(\lambda(t), \rho(t), \sigma, \mu\). These multipliers, labeled \(N = 0\), are shown in Table 1.

**Modified Quasilinearization Algorithm.** Together with (130)-(132), the optimal initial multipliers were employed as the nominal functions for the modified quasilinearization algorithm of Section 5. This algorithm was iterated until the stopping condition (116) was satisfied. Convergence to the solution was achieved in \(N = 6\) iterations.
The numerical results are presented in Tables 2-5. Table 2 shows the stepsizes $\alpha$, the cost functional $I$, the constraint error $P$, the error in the optimum conditions $Q$, and the cumulative error $R$ versus the iteration number $N$. Note that the minimum value of the cost functional for the constrained minimization problem is $I = 1.6561$. As expected, this value is higher than the minimum value of the cost functional $I = 1.5373$ for the unconstrained minimization problem (Ref. 19). Table 3 shows the actual entrance time $\theta_1$ and the actual exit time $\theta_2$ versus the iteration number $N$. Table 4 shows the functions $x(t)$, $u(t)$ for the converged solution $N = 6$. Finally, Table 5 shows the multipliers $\lambda(t)$, $\rho(t)$, $\sigma$, $\mu$ for the converged solution $N = 6$. 
Table 1. Initial multipliers for Example 7.1 (N = 0).

<table>
<thead>
<tr>
<th>t</th>
<th>$\lambda$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>-3.3930</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.2</td>
<td>-3.2223</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.4</td>
<td>-3.0614</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.6</td>
<td>-2.9096</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.8</td>
<td>-2.7666</td>
<td>0.0000</td>
</tr>
<tr>
<td>1.0</td>
<td>-2.6319</td>
<td>0.0000</td>
</tr>
<tr>
<td>1.0</td>
<td>-4.0574</td>
<td>-2.0562</td>
</tr>
<tr>
<td>1.2</td>
<td>-3.2399</td>
<td>-1.5598</td>
</tr>
<tr>
<td>1.4</td>
<td>-2.5111</td>
<td>-1.1251</td>
</tr>
<tr>
<td>1.6</td>
<td>-1.8631</td>
<td>-0.7446</td>
</tr>
<tr>
<td>1.8</td>
<td>-1.2883</td>
<td>-0.4106</td>
</tr>
<tr>
<td>2.0</td>
<td>-0.7789</td>
<td>-0.1152</td>
</tr>
<tr>
<td>2.0</td>
<td>-0.7789</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.2</td>
<td>-0.7153</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.4</td>
<td>-0.6527</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.6</td>
<td>-0.5910</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.8</td>
<td>-0.5304</td>
<td>0.0000</td>
</tr>
<tr>
<td>3.0</td>
<td>-0.4710</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

$\sigma = -1.4255$, $\mu = 0.4710$
Table 2. Results for Example 7.1.

<table>
<thead>
<tr>
<th>N</th>
<th>α</th>
<th>I</th>
<th>P</th>
<th>Q</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>1.4129</td>
<td>0.88E-01</td>
<td>0.26E-02</td>
<td>0.90E-01</td>
</tr>
<tr>
<td>1</td>
<td>1/2</td>
<td>1.5009</td>
<td>0.27E-01</td>
<td>0.68E-02</td>
<td>0.34E-01</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.6401</td>
<td>0.87E-03</td>
<td>0.61E-02</td>
<td>0.70E-02</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.6541</td>
<td>0.10E-04</td>
<td>0.85E-04</td>
<td>0.95E-04</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.6559</td>
<td>0.11E-06</td>
<td>0.14E-05</td>
<td>0.16E-05</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1.6561</td>
<td>0.47E-08</td>
<td>0.77E-07</td>
<td>0.82E-07</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1.6561</td>
<td>0.28E-09</td>
<td>0.47E-08</td>
<td>0.50E-08</td>
</tr>
</tbody>
</table>

Table 3. Results for Example 7.1.

<table>
<thead>
<tr>
<th>N</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1000</td>
<td>0.9000</td>
</tr>
<tr>
<td>1</td>
<td>0.1601</td>
<td>0.8506</td>
</tr>
<tr>
<td>2</td>
<td>0.2380</td>
<td>0.7757</td>
</tr>
<tr>
<td>3</td>
<td>0.2611</td>
<td>0.7396</td>
</tr>
<tr>
<td>4</td>
<td>0.2733</td>
<td>0.7261</td>
</tr>
<tr>
<td>5</td>
<td>0.2793</td>
<td>0.7203</td>
</tr>
<tr>
<td>6</td>
<td>0.2823</td>
<td>0.7175</td>
</tr>
</tbody>
</table>
Table 4. Converged state and control for Example 7.1 ($N = 6$).

<table>
<thead>
<tr>
<th>t</th>
<th>x</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>1.0000</td>
<td>1.7306</td>
</tr>
<tr>
<td>0.2</td>
<td>0.9633</td>
<td>1.4967</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9355</td>
<td>1.2938</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9159</td>
<td>1.1159</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9041</td>
<td>0.9582</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9000</td>
<td>0.8170</td>
</tr>
<tr>
<td>1.0</td>
<td>0.9000</td>
<td>0.8100</td>
</tr>
<tr>
<td>1.2</td>
<td>0.9000</td>
<td>0.8100</td>
</tr>
<tr>
<td>1.4</td>
<td>0.9000</td>
<td>0.8100</td>
</tr>
<tr>
<td>1.6</td>
<td>0.9000</td>
<td>0.8100</td>
</tr>
<tr>
<td>1.8</td>
<td>0.9000</td>
<td>0.8100</td>
</tr>
<tr>
<td>2.0</td>
<td>0.9000</td>
<td>0.8100</td>
</tr>
<tr>
<td>2.0</td>
<td>0.9000</td>
<td>0.8032</td>
</tr>
<tr>
<td>2.2</td>
<td>0.9041</td>
<td>0.6777</td>
</tr>
<tr>
<td>2.4</td>
<td>0.9159</td>
<td>0.5621</td>
</tr>
<tr>
<td>2.6</td>
<td>0.9355</td>
<td>0.4567</td>
</tr>
<tr>
<td>2.8</td>
<td>0.9633</td>
<td>0.3594</td>
</tr>
<tr>
<td>3.0</td>
<td>1.0000</td>
<td>0.2693</td>
</tr>
</tbody>
</table>
Table 5. Converged multipliers for Example 7.1 ($N = 6$).

<table>
<thead>
<tr>
<th>$t$</th>
<th>$\lambda$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>-3.4613</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.2</td>
<td>-2.9935</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.4</td>
<td>-2.5877</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.6</td>
<td>-2.2318</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.8</td>
<td>-1.9164</td>
<td>0.0000</td>
</tr>
<tr>
<td>1.0</td>
<td>-1.6341</td>
<td>0.0000</td>
</tr>
<tr>
<td>1.0</td>
<td>-3.6589</td>
<td>-2.0389</td>
</tr>
<tr>
<td>1.2</td>
<td>-3.2484</td>
<td>-1.6284</td>
</tr>
<tr>
<td>1.4</td>
<td>-2.8380</td>
<td>-1.2180</td>
</tr>
<tr>
<td>1.6</td>
<td>-2.4275</td>
<td>-0.8075</td>
</tr>
<tr>
<td>1.8</td>
<td>-2.0170</td>
<td>-0.3970</td>
</tr>
<tr>
<td>2.0</td>
<td>-1.6065</td>
<td>0.0134</td>
</tr>
<tr>
<td>2.0</td>
<td>-1.6065</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.2</td>
<td>-1.3541</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.4</td>
<td>-1.1242</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.6</td>
<td>-0.9134</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.8</td>
<td>-0.7189</td>
<td>0.0000</td>
</tr>
<tr>
<td>3.0</td>
<td>-0.5386</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

$\sigma = -2.0248$ , $\mu = 0.5386$
7.2. **Free-Final-Time Example.** Consider the problem of minimizing the cost functional

\[ I = \tau \]  

(133)

with respect to the state \( x(\theta), y(\theta) \), the control \( u(\theta) \), the parameter \( \tau \) which satisfy the differential constraints

\[ \frac{dx}{d\theta} = u \, , \quad \frac{dy}{d\theta} = x^2 - u^2 \, , \quad 0 \leq \theta \leq \tau \]  

(134)

the state variable inequality constraint

\[ y \geq -0.4 \, , \quad 0 \leq \theta \leq \tau \]  

(135)

and the boundary conditions

\[ x(0) = 0 \, , \quad x(\tau) = 1 \]  

(136)

\[ y(0) = 0 \, , \quad y(\tau) = 0 \]  

(137)

The solution obtained bypassing Ineq. (135) is given in Ref. 20 and violates Ineq. (135) over the interval \( 0.3 \leq \theta \leq 0.7 \). This being the case, we postulate that the extremal arc includes three subarcs, specifically,

\[ y \geq -0.4 \, , \quad 0 \leq \theta \leq \theta_1 \, \text{and} \, \theta_2 \leq \theta \leq \tau \]  

\[ y = -0.4 \, , \quad \theta_1 \leq \theta \leq \theta_2 \]  

(138)

When the normalized time \( t \) is employed, problem (133)-(137) can be reformulated as that of minimizing the cost functional

\[ I = \tau \]  

(139)
with respect to the state $x(t)$, $y(t)$, the control $u(t)$, and the parameter components $\theta_1$, $\theta_2$, $\tau$ which satisfy the differential constraints

$$
\dot{x} = \theta_1 u, \quad \dot{y} = \theta_1 (x^2 - u^2), \quad 0 \leq t \leq 1 \tag{140}
$$

$$
\dot{x} = (\theta_2 - \theta_1) u, \quad \dot{y} = (\theta_2 - \theta_1) (x^2 - u^2), \quad 1 \leq t \leq 2 \tag{141}
$$

$$
\dot{x} = (\tau - \theta_2) u, \quad \dot{y} = (\tau - \theta_2) (x^2 - u^2), \quad 2 \leq t \leq 3 \tag{142}
$$

the control variable equality constraint

$$
(\theta_2 - \theta_1) (x^2 - u^2) = 0, \quad 1 \leq t \leq 2 \tag{143}
$$

and the multipoint conditions

$$
x(0) = 0, \quad x(3) = 1 \tag{144}
$$

$$
y(0) = 0, \quad y(1) = -0.4, \quad y(3) = 0 \tag{145}
$$

Of course, implicit in the above formulation is the assumption that

$$
y \geq -0.4, \quad 0 \leq t \leq 1 \text{ and } 2 \leq t \leq 3 \tag{146}
y = -0.4, \quad 1 \leq t \leq 2
$$

In this problem,

$$
n = 2, \quad m = 1, \quad p = 3, \quad k = 1, \quad q = 2 \tag{147}
$$

Since $n + p + k + 1 = 7$, seven particular solutions are needed for each iteration.
We assume the nominal state

\[ x(t) = \frac{t}{3} , \quad y(t) = -0.4t , \quad 0 \leq t \leq 1 \]  \hspace{1cm} (148)

\[ x(t) = \frac{t}{3} , \quad y(t) = -0.4 , \quad 1 \leq t \leq 2 \]  \hspace{1cm} (149)

\[ x(t) = \frac{t}{3} , \quad y(t) = -1.2 + 0.4t , \quad 2 \leq t \leq 3 \]  \hspace{1cm} (150)

the nominal control

\[ u(t) = \cos(\pi t/6) , \quad 0 \leq t \leq 1 \text{ and } 2 \leq t \leq 3 \]  \hspace{1cm} (151)

\[ u(t) = x(t) , \quad 1 \leq t \leq 2 \]

and the nominal parameter components

\[ \theta_1 = \pi/6 , \quad \theta_2 = \pi/3 , \quad \tau = \pi/2 \]  \hspace{1cm} (152)

We observe that these nominal functions (labeled \( N = -1 \)) are consistent with
the multipoint conditions (144)-(145), with the state variable inequality
constraint (146), and with the control variable equality constraint (143).

**Auxiliary Minimization Algorithm.** Starting with the nominal functions
(148)-(152), we employed the auxiliary minimization algorithm of Section 6
in order to generate the optimal initial multipliers \( \lambda_1(t) , \lambda_2(t) , \rho(t) , \sigma , \mu_1 , \mu_2 \).
These multipliers, labeled \( N = 0 \), are shown in Table 6.

**Modified Quasilinearization Algorithm.** Together with (148)-(152), the
optimal initial multipliers were employed as the nominal functions for the
modified quasilinearization algorithm of Section 5. This algorithm was
iterated until the stopping condition (116) was satisfied. Convergence to the
solution was achieved in \( N = 6 \) iterations.

The numerical results are presented in Tables 7-10. Table 7 shows the stepsize \( \alpha \), the cost functional \( I \), the constraint error \( P \), the error in the optimum conditions \( Q \), and the cumulative error \( R \) versus the iteration number \( N \). Note that the minimum value of the cost functional for the constrained minimization problem is \( I = 1.5827 \). As expected, this value is higher than the minimum value of the cost function \( I = 1.5707 \) for the unconstrained minimization problem (Ref. 20). Table 8 shows the actual entrance time \( \theta_1 \), the actual exit time \( \theta_2 \), and the actual final time \( \tau \) versus the iteration number \( N \). Table 9 shows the functions \( x(t), y(t), u(t) \) for the converged solution \( N = 6 \). Finally, Table 10 shows the multipliers \( \lambda_1(t), \lambda_2(t), \rho(t), \sigma, \mu_1, \mu_2 \) for the converged solution \( N = 6 \).
Table 6. Initial multipliers for Example 7.2 (N = 0).

<table>
<thead>
<tr>
<th>t</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>2.1311</td>
<td>1.1995</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.2</td>
<td>2.1241</td>
<td>1.1996</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.4</td>
<td>2.1031</td>
<td>1.1999</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.6</td>
<td>2.0682</td>
<td>1.2003</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.8</td>
<td>2.0195</td>
<td>1.2006</td>
<td>0.0000</td>
</tr>
<tr>
<td>1.0</td>
<td>1.9574</td>
<td>1.2007</td>
<td>0.0000</td>
</tr>
<tr>
<td>1.0</td>
<td>2.0325</td>
<td>1.1377</td>
<td>-1.6958</td>
</tr>
<tr>
<td>1.2</td>
<td>1.8566</td>
<td>1.1377</td>
<td>-1.0347</td>
</tr>
<tr>
<td>1.4</td>
<td>1.6922</td>
<td>1.1377</td>
<td>-0.5858</td>
</tr>
<tr>
<td>1.6</td>
<td>1.5361</td>
<td>1.1377</td>
<td>-0.2674</td>
</tr>
<tr>
<td>1.8</td>
<td>1.3851</td>
<td>1.1377</td>
<td>-0.0344</td>
</tr>
<tr>
<td>2.0</td>
<td>1.2360</td>
<td>1.1377</td>
<td>0.1400</td>
</tr>
<tr>
<td>2.0</td>
<td>1.1867</td>
<td>1.1377</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.2</td>
<td>1.0109</td>
<td>1.1370</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.4</td>
<td>0.8213</td>
<td>1.1352</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.6</td>
<td>0.6183</td>
<td>1.1328</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.8</td>
<td>0.4020</td>
<td>1.1306</td>
<td>0.0000</td>
</tr>
<tr>
<td>3.0</td>
<td>0.1722</td>
<td>1.1296</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

$\sigma = -0.0629$, $\mu_1 = -0.1722$, $\mu_2 = -1.1296$
Table 7. Results for Example 7.2.

<table>
<thead>
<tr>
<th>N</th>
<th>( \alpha )</th>
<th>I</th>
<th>P</th>
<th>Q</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1.5707</td>
<td>0.10E+00</td>
<td>0.77E-01</td>
<td>0.18E+00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.5592</td>
<td>0.16E-02</td>
<td>0.19E-01</td>
<td>0.21E-01</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.5829</td>
<td>0.56E-04</td>
<td>0.85E-03</td>
<td>0.91E-03</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.5826</td>
<td>0.37E-05</td>
<td>0.23E-04</td>
<td>0.26E-04</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.5827</td>
<td>0.23E-06</td>
<td>0.68E-06</td>
<td>0.91E-06</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1.5827</td>
<td>0.14E-07</td>
<td>0.28E-07</td>
<td>0.43E-07</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1.5827</td>
<td>0.91E-09</td>
<td>0.18E-08</td>
<td>0.27E-08</td>
</tr>
</tbody>
</table>

Table 8. Results for Example 7.2.

<table>
<thead>
<tr>
<th>N</th>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
<th>( \tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5235</td>
<td>1.0471</td>
<td>1.5707</td>
</tr>
<tr>
<td>1</td>
<td>0.6156</td>
<td>0.9580</td>
<td>1.5592</td>
</tr>
<tr>
<td>2</td>
<td>0.6995</td>
<td>0.9334</td>
<td>1.5829</td>
</tr>
<tr>
<td>3</td>
<td>0.7426</td>
<td>0.9176</td>
<td>1.5826</td>
</tr>
<tr>
<td>4</td>
<td>0.7640</td>
<td>0.9101</td>
<td>1.5827</td>
</tr>
<tr>
<td>5</td>
<td>0.7742</td>
<td>0.9063</td>
<td>1.5827</td>
</tr>
<tr>
<td>6</td>
<td>0.7800</td>
<td>0.9044</td>
<td>1.5827</td>
</tr>
</tbody>
</table>
Table 9. Converged state and control for Example 7.2 (N = 6).

<table>
<thead>
<tr>
<th>t</th>
<th>x</th>
<th>y</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.8944</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1389</td>
<td>-0.1227</td>
<td>0.8835</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2745</td>
<td>-0.2337</td>
<td>0.8512</td>
</tr>
<tr>
<td>0.6</td>
<td>0.4035</td>
<td>-0.3221</td>
<td>0.7982</td>
</tr>
<tr>
<td>0.8</td>
<td>0.5226</td>
<td>-0.3791</td>
<td>0.7258</td>
</tr>
<tr>
<td>1.0</td>
<td>0.6290</td>
<td>-0.4000</td>
<td>0.6358</td>
</tr>
</tbody>
</table>

| 1.0 | 0.6290  | -0.4000 | 0.6290  |
| 1.2 | 0.6449  | -0.4000 | 0.6449  |
| 1.4 | 0.6611  | -0.4000 | 0.6611  |
| 0.6 | 0.6778  | -0.4000 | 0.6778  |
| 1.8 | 0.6948  | -0.4000 | 0.6948  |
| 2.0 | 0.7123  | -0.4000 | 0.7123  |

| 2.0 | 0.7123  | -0.4000 | 0.7096  |
| 2.2 | 0.8018  | -0.3809 | 0.6068  |
| 2.4 | 0.8765  | -0.3263 | 0.4927  |
| 2.6 | 0.9351  | -0.2401 | 0.3697  |
| 2.8 | 0.9765  | -0.1286 | 0.2398  |
| 3.0 | 1.0000  | 0.0000  | 0.1055  |
Table 10. Converged multipliers for Example 7.2 (N = 6).

<table>
<thead>
<tr>
<th>t</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>2.2360</td>
<td>1.2500</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.2</td>
<td>2.2089</td>
<td>1.2500</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.4</td>
<td>2.1281</td>
<td>1.2500</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.6</td>
<td>1.9956</td>
<td>1.2500</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.8</td>
<td>1.8146</td>
<td>1.2500</td>
<td>0.0000</td>
</tr>
<tr>
<td>1.0</td>
<td>1.5895</td>
<td>1.2500</td>
<td>0.0000</td>
</tr>
<tr>
<td>1.0</td>
<td>1.5895</td>
<td>0.9889</td>
<td>-0.2743</td>
</tr>
<tr>
<td>1.2</td>
<td>1.5505</td>
<td>0.9889</td>
<td>-0.2130</td>
</tr>
<tr>
<td>1.4</td>
<td>1.5124</td>
<td>0.9889</td>
<td>-0.1547</td>
</tr>
<tr>
<td>1.6</td>
<td>1.4753</td>
<td>0.9889</td>
<td>-0.0992</td>
</tr>
<tr>
<td>1.8</td>
<td>1.4390</td>
<td>0.9889</td>
<td>-0.0464</td>
</tr>
<tr>
<td>2.0</td>
<td>1.4037</td>
<td>0.9889</td>
<td>0.0037</td>
</tr>
<tr>
<td>2.0</td>
<td>1.4037</td>
<td>0.9889</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.2</td>
<td>1.2002</td>
<td>0.9889</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.4</td>
<td>0.9747</td>
<td>0.9889</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.6</td>
<td>0.7312</td>
<td>0.9889</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.8</td>
<td>0.4744</td>
<td>0.9889</td>
<td>0.0000</td>
</tr>
<tr>
<td>3.0</td>
<td>0.2088</td>
<td>0.9889</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

$\sigma = -0.2610$, $\mu_1 = -0.2088$, $\mu_2 = -0.9889$
8. **Remark**

The formulation in this thesis assumes that the functions $L_s(x, \theta)$ and $M_s(x, \theta)$ are linear in $x$ and $\theta$, so that the functions $L(x, \pi, t)$ and $M(x, \pi, t)$ are linear in $x$ and $\pi$. At first glance this appears to be a strong limitation to the present theory. Actually, this is not the case, because every constrained minimization problem can be brought to the present form through the introduction of some auxiliary state variables.

Assume that all the components of $M$ are nonlinear in $x$ and $\pi$. We can introduce $k$ additional state variables as follows:

$$y_1 = L(x, \pi, t), \quad y_2 = L_1(x, \pi, t), \quad \ldots, \quad y_k = L_{(k-1)}(x, \pi, t) \quad (153)$$

Then, we define the differential equations

$$\dot{y}_1 = L_1(x, \pi, t), \quad \dot{y}_2 = L_2(x, \pi, t), \quad \ldots, \quad \dot{y}_k = N(x, u, \pi, t) \quad (154)$$

subject to the initial conditions

$$y_1(0) = [L(x, \pi)]_0, \quad y_2(0) = [L_1(x, \pi)]_0, \quad \ldots, \quad y_k(0) = [L_{(k-1)}(x, \pi)]_0 \quad (155)$$

Now, the vector $M$ can be written as

$$M(y) = [y_1, y_2, \ldots, y_k]^T \quad (156)$$

Consequently, if all the components of $M$ are nonlinear in $x$ and $\pi$, $k$ additional state variables (153) are introduced, the differential system (20) is supplemented by the differential system (154), the initial condition (22-1) is supplemented
by (155), and (22-2) is replaced by

\[ [M(y)]_1 = 0 \]  \hspace{1cm} (157)

Clearly, if not all the components of M are nonlinear in x and \( \pi \), the dimension of the state vector is only increased by the number of the components that are nonlinear in x and \( \pi \).
9. Discussion and Conclusions

This thesis considers the numerical solution of optimal control problems involving a cost functional I subject to differential constraints, a state variable inequality constraint, and terminal constraints. The problem is to find the state $x(t)$, the control $u(t)$, and the parameter $\pi$ in such a way that the cost function is minimized, while the constraints are satisfied to a predetermined accuracy.

In contrast to penalty function techniques and transformation techniques, the state variable inequality constraint is handled in a direct manner. That is, for the time interval for which the trajectory of the system lies on the state boundary, the control vector is determined in such a way that the state boundary is satisfied to a predetermined accuracy.

To facilitate the numerical solution on digital computers, the actual time $\theta$ is replaced by the normalized time $t$, defined in such a way that each of the subarcs composing the extremal arc has a normalized time length $\Delta t = 1$. In this way, variable-time corner conditions and fixed-time terminal conditions are transformed into fixed-time-corner conditions and fixed-time terminal conditions. The actual time $\theta_1, \theta_2, \tau$ at which (i) the state boundary is entered, (ii) the state boundary is exited, and (iii) the terminal boundary is reached are regarded to be components of the parameter $\pi$ being optimized.

A modified quasilinearization algorithm is developed. Its main property is the descent property in the performance index $R$, the cumulative error in the constraints and the optimum conditions. Modified quasilinearization differs from ordinary quasilinearization because of the inclusion of the scaling factor.
(or stepsize) $\alpha$ in the system of variations. The stepsize is determined by a one-dimensional search on the performance index $R$. Since the first variation $\delta R$ is negative, the decrease in $R$ is guaranteed if $\alpha$ is sufficiently small. Convergence to the solution is achieved when $R$ becomes smaller than some preselected value.

In order to start the algorithm, some nominal functions $x(t), u(t), \pi$ and nominal multipliers $\lambda(t), \phi(t), \sigma, \mu$ must be chosen. In a real problem the selection of the nominal functions can be made on the basis of physical considerations. Concerning the nominal multipliers, no useful guidelines have been available thus far. In this paper, an auxiliary minimization algorithm for selecting the multipliers optimally is presented: the performance index $R$ is minimized with respect to $\lambda(t), \phi(t), \sigma, \mu$. Since the functional $R$ is quadratically dependent on the multipliers, the resulting variational problem is governed by optimality conditions which are linear, and therefore, can be solved without difficulty.

Two numerical examples are presented: the first is a fixed-final-time problem, and the second is a free-final-time problem. These examples demonstrate the feasibility as well as the rapidity of convergence of the technique developed in this thesis.

In closing, the following comments are pertinent.

(i) The updating technique employed in this thesis corresponds to the scheme used in Ref. 11. The functions $x(t), u(t), \pi$ and multipliers $\lambda(t), \phi(t), \sigma, \mu$ are updated in accordance with (40)-(41). This is called Scheme (a).

An alternate updating scheme is the following. One updates the control
with Eq. (40-2) and the parameter with Eq. (40-3). Then, the varied state
\( \bar{x}(t) \) is obtained by forward integration of Eq. (20) subject to the initial
condition (22-1). Concerning the multipliers, one updates \( \rho(t) \), \( \sigma \), \( \mu \) in
accordance with Eqs. (41-2), (41-3), (41-4). Then, the varied multiplier
\( \lambda(t) \) is obtained by backward integration of Eq. (28) subject to the multipoint
conditions (31). This is called Scheme (b).

It is conjectured that Scheme (a) is superior to Scheme (b) for problems
characterized by nonoscillatory solutions. On the other hand, the reverse
may be true for problems characterized by oscillatory solutions.

(ii) The analytical model considered in this thesis assumes that the
state variable boundary (23-2) is encountered only once. For this model, the
number of integrations required for each iteration is \( n + p + k + 1 \), where \( p = 2 \)
if \( \tau \) is fixed and \( p = 3 \) if \( \tau \) is free.

A more general case occurs if the state variable boundary (23-2) is
encountered \( s \) times. For this case, the number of integrations required for
each iteration is \( n + p + sk + 1 \), where \( p = 2s \) if \( \tau \) is fixed and \( p = 2s + 1 \) if
\( \tau \) is free.

(iii) The analytical model considered in this thesis assumes that the
left-hand side of the state variable inequality constraint \( L(x, \pi, t) \) is a linear
scalar function. Should \( L(x, \pi, t) \) be a linear vector function, the complexity
of the problem would increase. This is due to the fact that one would have
to guess not only the correct sequence of subarcs but, for those subarcs lying
on the state boundary, the active scalar components of the vector inequality
constraint.
(iv) If \( L(x, \pi, t) \) is a scalar function and \( s \) is large or if \( L(x, \pi, t) \) is a vector function, the technique outlined in this thesis may become computationally expensive, owing to the large number of integrations required for each iteration. For these cases, it is possible that penalty function techniques may prove to be more attractive computationally.
References


10. FONG, T.S., Method of Conjugate Gradient for Optimal Control Problems with State Variable Constraints, University of California at Los Angeles, School of Engineering and Applied Sciences, TR No. 70-30, 1970.


