IFAC

INTERNATIONAL FEDERATION
OF AUTOMATIC CONTROL

WARSZAWA 1969

Computational Methods in Optimisation

General Problems

Fourth Congress of the International Federation of Automatic Control
Warszawa 16–21 June 1969

TECHNICAL SESSION
18

NOT

Organized by
Naczelna Organizacja Techniczna w Polsce

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

Computational Methods in Optimisation

General Problems

TECHNICAL SESSION No 18

FOURTH CONGRESS OF THE INTERNATIONAL
FEDERATION OF AUTOMATIC CONTROL
WARSZAWA 16 – 21 JUNE 1969

Organized by
Naczelna Organizacja Techniczna w Polsce

Contents

Wydawnictwa Czasopism Technicznych NOT
Warszawa, ul. Czackiego 3/5 — Polska

Zakład Poligraficzny WCT NOT. Zam. 36/69.

DIFFERENTIAL DYNAMIC PROGRAMMING

D.H. JACOBSON* and D.Q. Mayne

Centre for Computing and Automation,
Imperial College, London.

1. Introduction

Consider the following control problem formulation:

$$\dot{x} = f(x,u;t) \; ; \; x(t_o) = x_o \tag{1}$$

$$V(x_o;t_o) = \int_{t_o}^{t_f} L(x,u;t)dt + F(x(t_f)) \tag{2}$$

$$g(u;t) \leqslant 0 \tag{3}$$

$$\psi(x(t_f)) = 0 \tag{4}$$

x is an n-dimensional state vector, u an m-dimensional control vector. f is an n-dimensional non-linear vector function, whilst L and F are non-linear scalar functions. g is a $p \leqslant m$-dimensional non-linear vector function and ψ is an $s \leqslant n$-dimensional non-linear vector function. The object of the control problem is to choose $u(t)$; $t \in [t_o, t_f]$ such that (3) and (4) are satisfied and V, given by (2), is minimized.

Merriam[1], Mitter[2] and McReynolds and Bryson[3] are some who have developed second-variation type algorithms for successively improving a nominal, guessed, control function. More recently Mayne[4] has developed a second-order algorithm using Dynamic Programming. (Differential Dynamic Programming) McReynolds[5] has, independently, obtained equivalent results. Jacobson[6,7,8,9,10] has further developed the notion of D.D.P. and has shown[7] that the second-variation algorithms[2,3] are only approximations to Mayne's second-order algorithm. Jacobson[6,8,9,10] has demonstrated that if instead of allowing only small changes in control at each iteration, large or global changes in control are permitted, then the D.D.P. algorithms which result are applicable to a much larger class of problems than existing algorithms. In particular, control inequality constrained problems can be solved. It is the purpose of this paper to report some of the new research results mentioned above, and to describe recent work on the application of D.D.P. to stochastic bang-bang problems.

* D.H. Jacobson is now with the Division of Engineering and Applied Physics, Harvard University, Cambridge, Mass. 02138, U.S.A.

2. Differential Dynamic Programming

Consider first the case where constraints (3) and (4) are absent. The optimal cost $V^o(x;t)$ satisfies the following P.D.E.

$$-\frac{\partial V^o}{\partial t}(x;t) = \min_u \left[L(x,u;t) + \langle V_x^o(x;t), f(x,u;t) \rangle \right] \tag{5}$$

Assume that the optimal contrul $u^o(t)$; $t \in [t_o, t_f]$ is unknown but that a nominal control $\bar{u}(t)$; $t \in [t_o, t_f]$ is available. On application of this nominal control, a nominal state trajectory $\bar{x}(t)$; $t \in [t_o, t_f]$ and nominal cost $\bar{V}(x_o;t_o)$ are produced by (1) and (2), respectively. Equations (1), (2) and (5) may be written in terms of these nominal values by setting

$$x = \bar{x} + \delta x, \quad u = \bar{u} + \delta u,$$
$$V(x_o;t_o) = \bar{V}(x_o;t_o) + a(x_o;t_o) \tag{6}$$

where δx and δu are the state and control variables, respectively, measured with respect to the nominal quantities $\bar{x}$, $\bar{u}$; they are not necessarily small quantities. $a(x_o;t_o)$ is the change in cost caused by using control $u = \bar{u} + \delta u$, instead of $u = \bar{u}$.

Equations (1), (2) and (5) become:

$$\frac{d}{dt}(\bar{x} + \delta x;t) = f(\bar{x} + \delta x, \bar{u} + \delta u;t) \; ; \; \bar{x}(t_o) + \delta x(t_o) = x_o \tag{7}$$

$$V(x_o;t_o) = \bar{V}(x_o;t_o) + a(x_o;t_o) = \int_{t_o}^{t_f} L(\bar{x} + \delta x, \bar{u} + \delta u;t)dt + F(\bar{x}(t_f) + \delta x(t_f)) \tag{8}$$

$$-\frac{\partial V^o}{\partial t}(\bar{x} + \delta x;t) = \min_{\delta u} \; L(\bar{x} + \delta x, \bar{u} + \delta u;t) + \langle V_x^o(\bar{x} + \delta x;t), f(\bar{x} + \delta x; \bar{u} + \delta u;t) \rangle \tag{9}$$

Assume now that the cost is smooth enough and δx is small enough to allow for a power series expansion for the cost, about $\bar{x}$, only up to second-order in δx: (Methods for keeping δx small are discussed later)

$$V(\bar{x} + \delta x;t) = V(\bar{x};t) + \langle V_x, \delta x \rangle + \tfrac{1}{2} \langle \delta x, V_{xx} \delta x \rangle \tag{10}$$

$$= \bar{V} + a + \langle V_x, \delta x \rangle + \tfrac{1}{2} \langle \delta x, V_{xx} \delta x \rangle \tag{11}$$

All quantities in (11) evaluated at $\bar{x};t$.

$$V_x(\bar{x} + \delta x;t) = V_x + V_{xx} \delta x \tag{12}$$

The subscript ^{o} on 'V quantities' in Equations (10)-(12) has been dropped for the following reason: Modelling the cost surface, locally, by a second-order expansion is made possible by keeping δx small. So the cost described by (11) is optimal subject to the proviso that δx remains small.

Substituting (11)-(12) into (9):

$$-\frac{\partial \bar{V}}{\partial t} - \frac{\partial a}{\partial t} - \left\langle \frac{\partial V_x}{\partial t}, \delta x \right\rangle - \tfrac{1}{2}\left\langle \delta x, \frac{\partial V_{xx}}{\partial t}\, \delta x \right\rangle =$$

$$\min_{\delta u}\left[L(x+\delta x, \bar{u}+\delta u; t) + V_x + V_{xx}\,\delta x,\ f(\bar{x}+\delta x, \bar{u}+\delta u; t)\right] \tag{13}$$

Now $V = (\bar{V} + a)$, V_x and V_{xx} are functions of x and t (14)

$$\frac{d}{dt}(V_x) = \dot{V}_x = \frac{\partial V_x}{\partial t} + V_{xx}\, f(\bar{x}, u; t)$$

$$\frac{d}{dt}(V_{xx}) = \dot{V}_{xx} = \frac{\partial V_{xx}}{\partial t}\ , \tag{15}$$

since higher order terms in V have been truncated. Also

$$-\dot{\bar{V}} = L(\bar{x}, \bar{u}; t) \tag{16}$$

Equations (13)-(16) will be used to develop algorithms for successively improving the nominal control $\bar{u}(t)$; $t \in [t_o, t_f]$.

3. Mayne's Second-Order Algorithm

Define $H(x,u,V_x;t) = L(x,u;t) + \langle V_x, f(x,u;t)\rangle$ (17)

Expanding the right hand side of (13), about $\bar{x}$, $\bar{u}$, to second-order in δx and δu and using (17), we obtain:

$$\min_{\delta u}\Big[H + \langle H_x + V_{xx} f, \delta x\rangle + \langle H_u, \delta u\rangle + \langle \delta u, (H_{ux} + f_u^T V_{xx})\delta x\rangle$$

$$+ \tfrac{1}{2}\langle \delta u, H_{uu}\,\delta u\rangle + \tfrac{1}{2}\langle \delta x, (H_{xx} + f_x^T V_{xx} + V_{xx} f_x)\delta x\rangle\Big] \tag{18}$$

In order to minimize, assuming $H_{uu}(\bar{x}, \bar{u}, V_x; t)$ to be positive definite, we differentiate with respect to δu and equate to zero; this yields:

$$\delta u = -H_{uu}^{-1}\left[H_u + (f_u^T V_{xx} + H_{ux})\,\delta x\right] \tag{19}$$

δu given by (19) may be so large that it invalidates the expansion, to second-order, in (18), so ε ; $0 < \varepsilon \leq 1$ is introduced to limit the size of δu and δx. Equation (19) becomes:

$$\delta u = -H_{uu}^{-1}\left[\varepsilon H_u + (H_{ux} + f_u^T V_{xx})\delta x \right] \tag{20}$$

If (20) is substituted back into (18), an expression consisting of constant terms, terms in δx, and terms in δx^2, is obtained. Since the left hand side of (13) equals the expression for all δx, coefficients of like powers of δx may be equated to obtain expressions for

$$-\frac{\partial \bar{V}}{\partial t} - \frac{\partial a}{\partial t}, \quad -\frac{\partial V_x}{\partial t} \quad \text{and} \quad -\frac{\partial V_{xx}}{\partial t} \quad .$$

Using these expressions and equations (14)-(16), the following differential equations are obtained:

$$-\dot{a} = -\varepsilon\left(1 - \frac{\varepsilon}{2}\right) \langle H_u, H_{uu}^{-1} H_u \rangle$$

$$-\dot{V}_x = H_x + \beta^T H_u \tag{21}$$

$$-\dot{V}_{xx} = H_{xx} + f_x^T V_{xx} + V_{xx} f_x - (H_{ux} + f_u^T V_{xx})^T H_{uu}^{-1} (H_{ux} + f_u^T V_{xx})$$

where $\beta = -H_{uu}^{-1} (H_{ux} + f_u^T V_{xx})$ (22)

These differential equations may be integrated backwards from t_f using boundary conditions:

$$a(t_f) = 0, \quad V_x(t_f) = F_x(\bar{x}(t_f)), \quad V_{xx}(t_f) = F_{xx}(\bar{x}(t_f)) \tag{23}$$

If the new control $u(t) = \bar{u}(t) + \delta u(t)$ is applied to the system, then a reduction in cost is obtained, for ε sufficiently small. ($a(x_o;t)$, the predicted change in cost, is negative if H_{uu}^{-1} is positive-definite. For ε sufficiently small, the second-order expansions are valid.)

The above algorithm requires the integration of n less differential equations than the second-variation method. Jacobson[7], using a detailed analysis, has shown that the second-variation methods are only approximations to Mayne's algorithm; the D.D.P. algorithm is a simpler, more accurate method.

A serious disadvantage of all these methods is that $H_{uu}^{-1}(\bar{x},\bar{u},V_x;t)$ must be positive-definite. The next section describes an algorithm which does not suffer from this drawback.

4. Jacobson's Second-Order Algorithm[8]

Considering δx set equal to zero in (13), i.e. consider state $x = \bar{x}$ at time t.

$$-\frac{\partial \bar{V}}{\partial t} - \frac{\partial a}{\partial t} = \min_{\delta u} H(\bar{x}, \bar{u} + \delta u, V_x; t) \tag{24}$$

Now, instead of making small changes in control, let us minimize H with respect to δu. (The minimization may be done analytically or, if necessary, numerically.)

$$u^* = \bar{u} + \delta u^* \tag{25}$$

Equation (24) becomes:

$$-\frac{\partial \bar{V}}{\partial t} - \frac{\partial a}{\partial t} = H(\bar{x}, u^*, V_x; t) \tag{26}$$

Now we re-introduce the variation δx in x:

$$-\frac{\partial \bar{V}}{\partial t} - \frac{\partial a}{\partial t} - \left\langle \frac{\partial V_x}{\partial t}, \delta x \right\rangle - \frac{1}{2}\left\langle \delta x, \frac{\partial V_{xx}}{\partial t}\, \delta x \right\rangle$$

$$= \min_{\delta u} \left[H(\bar{x} + \delta x,\ u^{*} + \delta u,\ V_x; t) + \left\langle V_{xx}\, \delta x,\ f(\bar{x} + \delta x,\ u^{*} + \delta u; t) \right\rangle \right] \qquad (27)$$

δu is now measured with respect to u^{*}.

Expanding the right hand side of (27), about $\bar{x}, u^{*}$, the same expression as (18) is obtained except that all quantities are evaluated at $\bar{x}, u^{*}$ and not $\bar{x}, \bar{u}$. The equation relating δu to δx is found to be:

$$\delta u = -H_{uu}^{-1}\,(H_{ux} + f_u^{T} V_{xx})\delta x = \beta \delta x \qquad (28)$$

($H_u(\bar{x}, u^{*}, V_x; t) = 0$ because u^{*} minimizes $H(\bar{x}, u^{*}, V_x; t)$)

All quantities evaluated at $\bar{x}, u^{*}$.

Equation (28) requires that $H_{uu}^{-1}(\bar{x}, u^{*}, V_x; t)$ be positive-definite; this is far less restrictive than requiring $H_{uu}^{-1}(\bar{x}, \bar{u}, V_x; t)$ to be positive-definite as in Section 3.

Employing the same approach as described in Section 3, the following differential equations are obtained:

$$\begin{aligned} -\dot{a} &= H - H(\bar{x}, \bar{u}, V_x; t) \\ -\dot{V}_x &= H_x + V_{xx}(f - f(\bar{x}, \bar{u}; t)) \\ -\dot{V}_{xx} &= H_{xx} + f_x^{T} V_{xx} + V_{xx} f_x \\ &\quad - (H_{ux} + f_u^{T} V_{xx})^{T} H_{uu}^{-1} (H_{ux} + f_u^{T} V_{xx}) \end{aligned} \qquad (29)$$

Unless otherwise stated, all quantities are evaluated at $\bar{x}, u^{*}$.

The new control to be applied is given by:

$$u = \bar{u} + \delta u^{*} + \beta \delta x = u^{*} + \beta \delta x \qquad (30)$$

This control may produce δx's which are 'too large', owing to the fact that δu^{*} might be very large. One cannot, however, place ε ; $0 < \varepsilon \leqslant 1$ in front of δu^{*} as was done in Section 3, to limit the size of δx, because $u^{*} = \bar{u} + \delta u^{*}$ is imbedded in the reverse differential equations (29) which have already been integrated. The next section describes a new method for limiting the size of δx.

5. A New 'Step-Size Adjustment Method'[8]

This method is described in detail elsewhere[8]; here, only a brief description is given.

When integrating the reverse differential equations (29), note the time t_{eff} at which $a(\bar{x};t)$ becomes different from zero ($a(\bar{x};t) = 0$ is a necessary condition of optimality. $a(\bar{x};t)$ is always ≤ 0 because u^* minimizes H.)

On the forwards run, try applying $u = u^* + \beta\delta x$ for $t \in [t_o, t_f]$. If a reduction in cost is obtained, then proceed to the next iteration. If no reduction in cost is obtained, run along the old nominal $\bar{x}(t)$ trajectory for a time $[t_o, t_1]$ where:

$$t_1 = \frac{t_{eff} - t_o}{2} + t_o = t_{o1} \tag{31}$$

and then apply $u = u^* + \beta\delta x$ on the time interval $[t_1, t_f]$. If still no reduction in cost is obtained, set:

$$t_1 = \frac{t_{eff} - t_{o1}}{2} + t_{o1} = t_{o2} \tag{32}$$

and repeat. In general:

$$t_1 = \frac{t_{eff} - t_{or}}{2} + t_{or} = t_{or+1} \tag{33}$$

where $r = 0, 1 \ldots$ and $t_{oo} = t_o$ (34)

Ultimately a t_1, sufficiently close to t_{eff}, will be found such that the δx's produced on the time interval $[t_1, t_f]$ will be small enough, and a reduction in cost will be achieved. (The δx's produced on the time interval $[t_1, t_f]$ are due to the fact that $u^*(t) \neq \bar{u}(t)$; $t \in [t_1, t_{eff}]$. For $t \in [t_{eff}, t_f]$, $u^*(t) = \bar{u}(t)$.) As optimality is approached, so $t_{eff} \rightarrow t_o$.

6. Illustrative Example

Consider the scalar control problem:

$$\dot{x} = -\cdot 2x + 10 \tanh u \quad ; \quad x(t_o) = 5 \tag{35}$$

Choose $u(t)$; $t \in [0, \cdot 5]$ to minimize:

$$V = \int_0^{\cdot 5} (10x^2 + u^2)dt + 10x^2(t_f) \tag{36}$$

The problem, though simple, is a good illustrative one because, along certain non-optimal trajectories $H_{uu}^{-1}(\bar{x},\bar{u},V_x;t)$ is not positive.

$$H(x,u,V_x;t) = 10x^2 + u^2 + V_x(-\cdot 2x + 10 \tanh u) \tag{37}$$

$$H_u = 2u + 10 V_x (1 - \tanh^2 u) \tag{38}$$

$$H_{uu} = 2 - 20\, V_x \,\text{Tanh}\, u\, (1 - \text{Tanh}^2 u) \qquad (39)$$

It is clear that there is no guarantee that H_{uu} will be positive for all nominal trajectories $\bar{x},\bar{u}$, i.e. there is no guarantee that the method of Section 3 or the second-variation methods will be successful.

At $u = u^*$, we have, from (38), that:

$$1 - \text{Tanh}^2 u^* = -\frac{2u^*}{10V_x} \quad ; \quad V_x \neq 0$$

$$= 1 \quad ; \quad V_x = 0 \qquad (40)$$

Using (40) in (39):

$$H_{uu}(\bar{x},u^*,V_x;t) = 2 + 4u^* \,\text{Tanh}\, u^* \qquad (41)$$

From (41), since $u^* \,\text{Tanh}\, u^* \geqslant 0$ for all u^*, $H_{uu}(\bar{x},u^*,V_x;t) > 0$ regardless of the nominal trajectory. The new algorithm if Section 4 should, therefore, not fail to solve this problem.

Using a nominal control:

$$\bar{u}(t) = +1 \quad ; \quad t \in [0,\ .5] \qquad (42)$$

$H_{uu}(\bar{x},\bar{u},V_x;t)$ turned out to be negative on the whole time interval and so the algorithm of Section 3, and the second-variation methods, were unable to improve the trajectory. Starting from the same nominal, the new algorithm converged to the optimal trajectory (optimal cost 41.6) in 4 iterations. A fourth-order Runge-Kutta integration routine was used. The interval $[0,\ .5]$ was divided into 100 steps. Fig. 1 shows the nominal and optimal controls. Further computational experience with the new algorithm is described elsewhere[8,10].

7. Constrained Problems

Control problems with constraints (3) and (4) have been considered[6,8,9,10] and the algorithm of Section 4 has been generalised to solve these problems. In the next section an algorithm for a special class of constrained problems is described.

8. A New Second-Order Algorithm for Solving Bang-Bang Control, Free End-Point Problems[8]

Consider the control problem formulation of Section 1, where:

$$f(x,u;t) = f_1(x;t) + f_2(x;t)u$$

$$L(x,u;t) = L(x;t) \qquad (43)$$

$$g(u;t) \text{ is of form } u^a \leqslant u \leqslant u^b$$

u is a scalar control variable (the algorithm is extendible to the case of a vector control variable[9].) and u^a and u^b are constants.

Constraints (4) are assumed to be absent. (End-Point constrained bang-bang control problems are treated in Refs.[6,9,10].)

It is easy to show[6,9,10] that, between switch points of the control u^*, the following differential equations hold:

$$\begin{aligned} -\dot{a} &= H - H(\bar{x},\bar{u},V_x;t) \\ -\dot{V}_x &= H_x + V_{xx}(f - f(\bar{x},\bar{u};t)) \\ -\dot{V}_{xx} &= H_{xx} + f_x^T V_{xx} + V_{xx} f_x \end{aligned} \tag{44}$$

All quantities evaluated at $\bar{x},u^*$.

u is given by:

$$\begin{aligned} u^* &= u^a \quad \text{if} \quad f_2^T V_x > 0 \\ &= u^b \quad \text{if} \quad f_2^T V_x < 0 \end{aligned} \tag{45}$$

The equations (44) can be integrated backwards from the usual boundary conditions, (23), until a switch point of u^*, t_s say, is reached.

At $t = t_s$, it is found that:

1) a is continuous across a switch point
2) V_x " " " " " "
3) V_{xx} experiences a jump, at t_x, given by:

$$\Delta V_{xx} = - V_{xt_s} \cdot V_{t_s x} / V_{t_s t_s} \tag{46}$$

where $V_{xt_s} = H_x^- - H_x^+ + V_{xx}(f^- - f^+)$

$$V_{t_s t_s} = H_t^- - H_t^+ - \langle H_x^-, (f^- - f^+) \rangle \tag{47}$$

$$+ \langle f^-, (H_x^- - H_x^+) \rangle + \langle f^- - f^+, V_{xx}(f^- - f^+) \rangle$$

(Recently, and independently, McReynolds[11] has obtained similar results.) Subscripts $^-$ and $^+$ denote that the quantities are evaluated at the control u^* just before, and just after the switch point t_s, respectively.

4) Variations in the switch time t_s are related to variations in the state x, by the linear feedback controller:

$$\delta t_s = \langle \beta, \delta x \rangle \tag{48}$$

where

$$\beta = - V_{t_s t_s}^{-1} V_{xt_s} \tag{49}$$

8.1 Computational Procedure

Equations (44) are integrated backwards in time until a switch point t_s is reached; here, V_{xx} is calculated, and the integration of (44) is resumed, using boundary conditions at t_s of:

$$V_{xx}(t_s) = V_{xx}(t_s^+) + \Delta V_{xx} \tag{50}$$

where superscript $^+$ in (50) denotes time immediately after t_s.

Any finite number of switch times, n_s, may be handled in this way:

$$\delta t_{s_i} = \langle \beta_i, \delta x(t_{s_i}) \rangle \; ; \; i = 1 \ldots n_s \tag{51}$$

The new control function, which yields a new improved trajectory, is calculated in the following way:

The new control is set as:

$$u(t) = u^*(t) \tag{52}$$

When a switch time t_{s_i} of u^* is reached, measure $\delta x(t_{s_i})$ and calculate δt_{s_i} using (51). if $\delta t_{s_i} > 0$, hold $u(t) = u^*(t_{s_i}^-)$ for the time interval δt_{s_i} ; after this, once again use (52) and continue. If, however, $\delta t_{s_i} < 0$, then back-space the integration routine by the amount δt_{s_i} and, starting at this time $t_{s_i} - \delta t_{s_i}$, set $u(t) = u^*(t_{s_i}^+)$ for $t_{s_i} - \delta t_{s_i} \leq t \leq t_{s_i}$ and integrate forwards again. When t_{s_i} is reached once again use (52) and continue.

The above procedure implements the local, linear feedback controller (51), directly. Applying the new control function on the whole time interval $[t_o, t_f]$ may produce δx's which are too large (i.e. a reduction in cost is not achieved). The 'step-size adjustment method' of Section 5 must then be used to limit the size of the δx's.

9. Illustrative Example

Consider the following 4th order control system[12].

$$\begin{aligned}
\dot{x}_1 &= -0.5x_1 + 5x_2 \quad ; \quad x_1(0) = 10 \\
\dot{x}_2 &= -5x_1 - 0.5x_2 + u \; ; \quad x_2(0) = 10 \\
\dot{x}_3 &= -0.6x_3 + 10x_4 \quad ; \quad x_3(0) = 10 \\
\dot{x}_4 &= -10x_3 - 0.6x_4 + u \; ; \quad x_4(0) = 10
\end{aligned} \tag{53}$$

$$|u| \leq 1$$

Minimize $V = \langle x(t_f), x(t_f) \rangle$; $t_f = 4.2$ secs. (55)

A Runge-Kutta, fourth-order integration routine was used. The time interval $[0, 4.2]$ was divided into 300 steps. A nominal control

$$\bar{u}(t) = +1 \quad ; \quad t \in [0, 4.2] \tag{56}$$

was used and this produced a cost of 4.12 units. The second-order

algorithm reduced this cost to the minimum value of .996 in two iterations. Figure 2 shows the cost as a function of iteration number, and Figure 3 shows the nominal, optimal controls. Note the marked difference between these controls.

10. Stochastic Control Problems

A feature of D.D.P. is that it yields an approximation to the cost curve $V(\bar{x}, t)$ in the neighbourhood of $\bar{x}$ whatever the subsequent policy π (defined by $u(\tau) = h(x;\tau)$, $\tau \in [t,t_f]$) is. Let $f'(x;\tau) = f(x,h(x;\tau);\tau)$ and let $\bar{x}(\tau)$ denote the solution of equation (1) with initial condition $\bar{x}(t)$ at time t in response to a nominal control $\bar{u}(\tau)$, $\tau \in [t,t_f]$. Then, if either $[t,t_f]$ is sufficiently small or $u(\tau) - \bar{u}(\tau)$, $\tau \in [t,t_f]$, is sufficiently small, the parameters $a(t)$, $V_x(t)$, $V_{xx}(t)$ which define $V(x;t)$ in the neighbourhood of $\bar{x}$ (equation (11)) are solutions of the following differential equations:

$$-\dot{a} = H' - H(\bar{x},\bar{u},V_x;t) \tag{57}$$

$$-\dot{V}_x = H_x' + V_{xx}(f' - f(\bar{x},\bar{u};t)) \tag{58}$$

$$-\dot{V}_{xx} = H_{xx}' + (f_x')^T V_{xx} + V_{xx} f_x' \tag{59}$$

with boundary conditions given in equation (23).

The unspecified arguments are $\bar{x}$ and t, and

$$H'(x,V_x;t) \triangleq H(x,h(x;t),V_x;t) \tag{60}$$

Consider now a stochastic system:

$$\dot{x} = f(x,u;t) + w, \quad x(t_o) = x_o \tag{61}$$

where w is piecewise constant (over intervals $\Delta = \frac{t_f - t_o}{N}$)

$$w(t) = w_i, \quad i\Delta \leq t < (i+1)\Delta \tag{62}$$

and $\omega = w_1 \ldots w_N$ is a sequence of independent random variables (simulation of a stochastic system necessitates a 'physical' disturbance). The cost of an individual realisation ($\omega = \omega_j$, say) with initial condition $x(t)$ and policy π is:

$$V(\bar{x},\omega_j;t) = \int_t^{t_f} L(x,u,t)dt + F(x(t_f)) \tag{63}$$

where $x(\tau) = x(\omega_j;\tau)$ is the solution at τ of equation (61), with initial condition $x(t) = \bar{x}$, policy π and w specified by ω_j. The cost of the process with initial condition $x(t) = \bar{x}$ is:

$$V_a(\bar{x};t) = \underset{\omega}{E}\, V(\bar{x},\omega;t) \tag{64}$$

($E_\omega(\cdot)$ denote expectation with respect to the distribution of ω) and the average cost of the process over $[t,t_f]$ is

$$\phi(t) = \underset{x(t)}{E} \; V_a(x(t);t) \tag{65}$$

Equations (57)-(59) may be integrated along each realisation to yield the random variables $V_x(\bar{x},\omega;t)$ and $V_{xx}(\bar{x},\omega;t)$. Under certain regularity conditions:

$$\frac{\partial V_a(\bar{x};t)}{\partial t} = \underset{|x(t)=\bar{x}}{E}\left[\frac{\partial V(\bar{x},\omega;t)}{\partial t}\right] \tag{66}$$

Using the procedure of Section 4 yields:

$$\begin{aligned} -\frac{\partial V_a(\bar{x};t)}{\partial t} = \underset{|x(t)=\bar{x}}{E}\Big[& H - H(\bar{x},\bar{u},V_x;t) \\ & + \delta x^T V_{xx}(f - f(\bar{x},\bar{u};t)) \\ & + H_x^T \delta x + H_u^T \delta u + \tfrac{1}{2}\delta x^T H_{xx}\delta x \\ & + \delta u^T H_{ux}\delta x + \tfrac{1}{2}\delta u^T H_{uu}\delta u \\ & + \delta x^T V_{xx}(f_x \delta x + f_u \delta u)\Big] \end{aligned} \tag{67}$$

where the unspecified arguments are $\bar{x}$, u^*, t, and $\delta x = x - \bar{x}$, $\delta u = u - u^*$, and u^* minimises $\underset{|x(t)=\bar{x}}{E} H(\bar{x},u,V_x;t)$. Hence:

(1) The control u^* at t minimises the cost of the process given $x(t) = \bar{x}$ and policy π for $\tau \in (t,t_f]$. Also $\underset{|x(t)=\bar{x}}{E} H_u(\bar{x},u^*,V_x;t) = 0$.

(2) The control $u^* + \beta\delta x$ at t minimises the cost of the process given $x(t) = \bar{x} + \delta x$ and policy π for $\tau \in (t,t_f]$ where:

$$\beta = -\left[\underset{|x(t)=\bar{x}}{E}(H_{uu})\right]^{-1}\left[\underset{|x(t)=\bar{x}}{E}(H_{ux} + f_u^T V_{xx})\right] \tag{68}$$

(3) Let u^o minimise $E\,H(x(t),u,V_x;t)$. Then u^o minimises $\phi(t)$, given policy π for $\tau \in (t,t_f]$; i.e. u^o is the optimal open loop control at t. Also $E\,H_u(x(t), u^o, V_x;t) = 0$.

As a corollary the optimal α, β in the parametrised control law $u(t) = \alpha(t) + \beta(t)\,x(t)$ satisfy:

$$E\,H_u(x(t),u^o,V_x;t) = 0 \quad , \quad E\,H_u(x(t),u^o V_x;t)\,x^T(t) = 0$$

where $u^o = \alpha^o + \beta^o x(t)$. $u(t)$ is a scalar.

Naive Monte Carlo simulation may be used to estimate the various quantities required to obtain $u^*, \beta, u^o, \alpha^o, \beta^o$ etc. The antithetic variate

method (applied first to control problems in Ref. 13) and the control variate method[14] (using a linear model obtained by statistical linearisation) may be used to reduce sampling variance. In fact it can be shown (if $p(\omega)$ is symmetric and $h(x;\tau)$ is linear in x, $\tau \in (t,t_f]$) that:

(1) For the L.Q.P. problem naive Monte Carlo yields zero variance estimates of β

(2) For the L.Q.P. problem the antithetic variate method yields zero variance estimates of $\underset{|x(t)=\bar{x}}{E}[V_x]$, $\underset{|x(t)=\bar{x}}{E}[H]$, $\underset{|x(t)=\bar{x}}{E}[H_u]$, $E[V_x]$, $E[H_u]$, and hence zero variance estimates of u^* and u^o

(3) For the L.Q.P. problem the control variate method yields zero sampling variance for all the above quantities and also quantities such as $E[H_u x^T]$, and therefore, α^o, β^o.

The consequential algorithm for determining optimal open loop control (or parametrised feedback control) is similar to the deterministic algorithm, by J trajectories (realisation), given $x(t_o) = x_o$, are used. Zero sampling variance for estimating $E[V_{t_s}]$ and $E[V_{t_s t_s}]$, required for optimal open-loop bang-bang control, can also be obtained for the L.Q.P. time-invariant problem using the antithetic variate method.

References

1. MERRIAM, C.W. III, 'Optimization Theory and the Design of Feedback Controls', McGraw Hill, 1964.
2. MITTER, S.K., 'Function Space Method in Optimal Control with Application to Power Systems', Ph.D. Thesis, Univ. of London, 1965
3. McREYNOLDS, S.R. and BRYSON, A.E., 'A Successive Sweep Method for Solving Optimal Programming Problems', J.A.C.C., No. 6, 1965, p. 551.
4. MAYNE, D.Q., 'A Second-Order Gradient Method of Optimizing Non-Linear Discrete Time Systems', International Journal of Control, Vol. 3, No. 1, 1966
5. McREYNOLDS, S.R., 'The Successive Sweep Method and Dynamic Programming', J. Math. Anal. Appl., Vol. 19, No. 3, 1967, pp. 565-598
6. JACOBSON, D.H., 'Differential Dynamic Programming Methods for Determining Optimal Control of Non-Linear Systems', Ph.D. Thesis, Univ. of London, 1967

7. JACOBSON, D.H., 'Second-Order and Second-Variation Methods for Determining Optimal Control: A Comparative Study Using Differential Dynamic Programming', International Journal of Control, 7, 1968, pp. 175-196

8. JACOBSON, D.H., 'New Second-Order and First-Order Algorithms for Determining Optimal Control: A Differential Dynamic Programming Approach', Journal of Optimization Theory and Applications, Vol. 2, No. 6, 1968 - To appear. Also, Harvard Technical Report, No. 551, February 1968

9. JACOBSON, D.H., 'Differential Dynamic Programming Methods for solving Bang-Bang Control Problems' - To appear

10. JACOBSON, D.H. and MAYNE, D.Q., 'Differential Dynamic Programming Book to appear, 1969

11. DYER, P. and McREYNOLDS, S.R., 'On Optimal Control Problems with Discontinuities', J. Math. Anal. Appl. - To appear

12. PLANT, J.B. and ATHANS, M., 'An Iterative Technique for the Computation of Time Optimal Controls', I.F.A.C., 3, London, 1966

13. MAYNE, D.Q., 'A Gradient Method for Determining Optimal Control of Non-Linear Stochastic Systems', Proc. I.F.A.C. Symposium on Adaptive Control, Teddington, 1965

14. WESTCOTT, J.H., MAYNE, D.Q., BRYANT, G.F. and MITTER, S.K., 'Optimal Techniques for On-Line Control', Proc. Third I.F.A.C. Congress, London, 1966

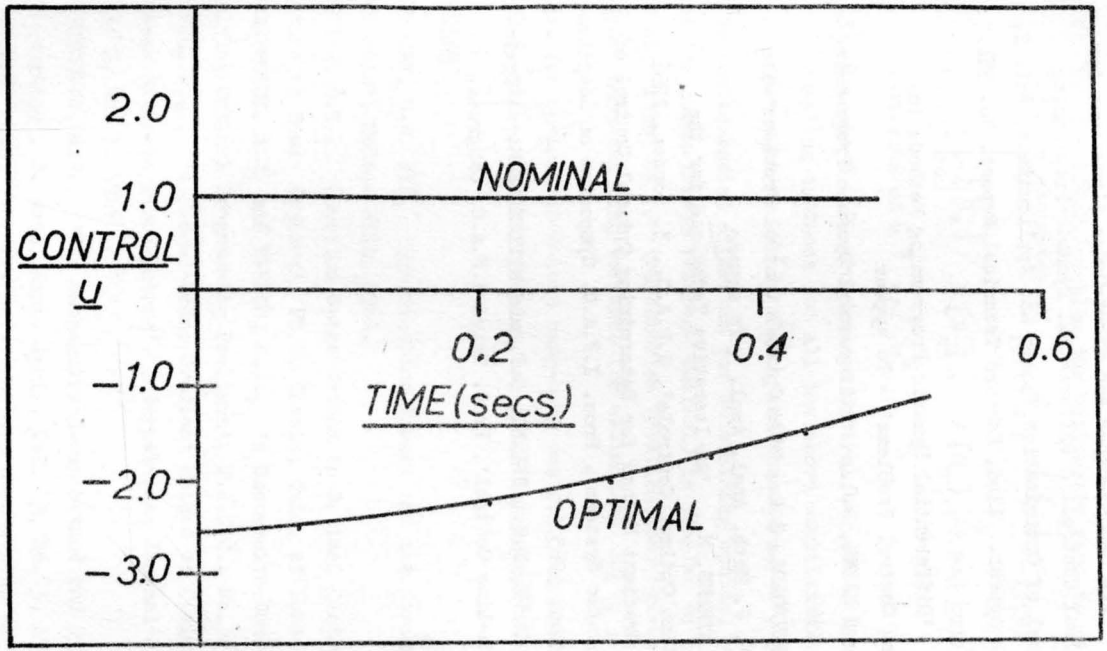


Fig. 1 The Scalar Control Problem: Nominal and Optimal Control Functions

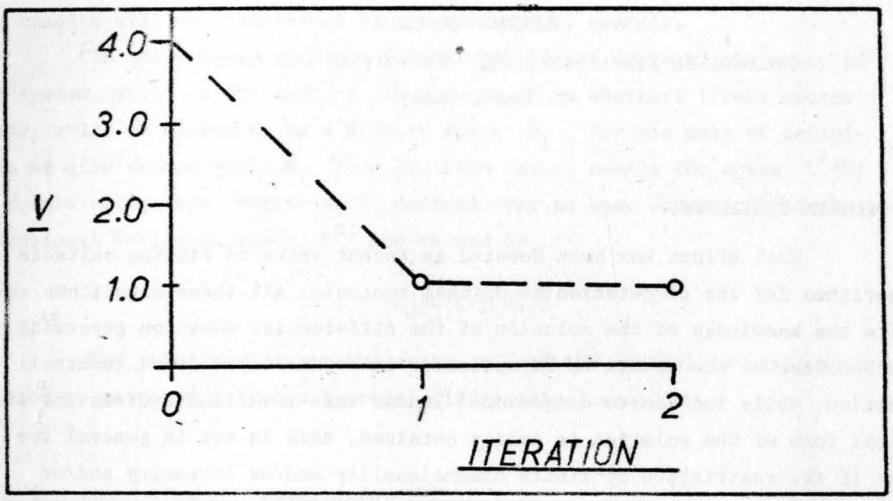


Fig. 2 Example II: Cost vs Iteration

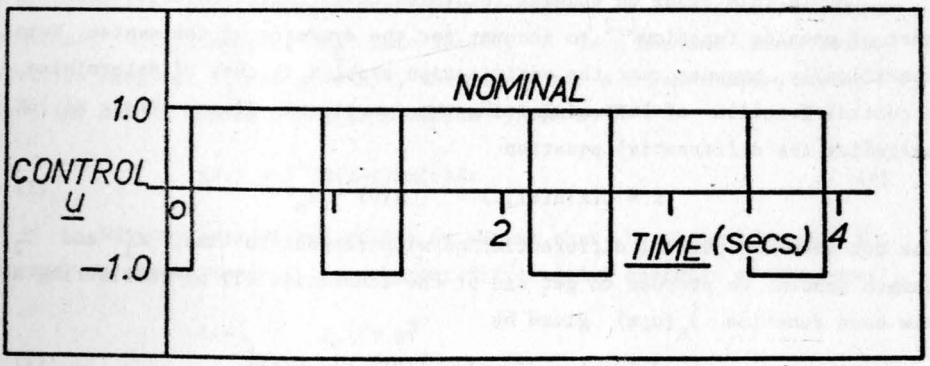


Fig. 3 Example II: Nominal and Optimal Controls

THE OPTIMIZATION OF DYNAMICAL SYSTEMS

SERGIO DE JULIO
Istituto di Elettrotecnica, Universita' di Roma
Rome, Italy

1. PROBLEM STATEMENT.

Much effort has been devoted in recent years to finding suitable algorithms for the computation of optimal controls. All these algorithms require the knowledge of the solution of the differential equation governing the behavior of the system to be optimized in terms of the input (control) function. While for finite-dimensional linear time-invariant systems the explicit form of the solution is easily obtained, this is not in general the case if the restriction of finite dimensionality and/or linearity and/or time invariance are relaxed.

In the case where an explicit expression for the solution of the system equation is not available and an iterative method of solution for the optimization problem is being used, at each iteration step there is the need for the computer to solve the differential equation of the system. As a result the computer memory requirements as well as the amount of computing time may turn out to be prohibitive.

In this paper we suggest to bypass these difficulties by using a sort of penalty function[1,2] to account for the dynamics of the system. More specifically, suppose that the optimization problem is that of determining a control function $u(\cdot)\epsilon X_1$ minimizing the functional $J(u;x)$ where $x(\cdot)\epsilon X_2$ satisfies the differential equation

$$\dot{x} = f(x;u(t);t) \qquad x(0) = x_o \tag{1}$$

the dot denoting partial differentiation with respect to time, X_1 and X_2 Banach spaces. We propose to get rid of the constraint (1) by considering a new cost function $J_\varepsilon(u;x)$ given by

$$J_\varepsilon(u;x) = J(u;x) + \frac{1}{\varepsilon} P(\dot{x}-f(x;u;t)) \tag{2}$$

and to carry out the minimization of J_ε over a suitable domain. To this new problem we shall refer to as the ε-problem.

We shall show that by appropriately choosing the functional P the solution of the ε-problem exists and can be made arbitrarily close to the solution of the original problem with an ε sufficiently small.

In this work we shall confine ourselves to the study of infinite

dimensional linear systems using the convenient framework of Hilbert spaces. The results will be illustrated by a computational example.

In the case of infinite dimensional linear systems, the state of the system $x(t)$ is for each t an element of an abstract linear vector space, which we assume to be a Hilbert space H_1. For the sake of definition we also assume that H_1 is a function space, namely the space $L^2(\Omega)$ of square integrable functions f defined over an open domain Ω of the n-dimensional Euclidean space R^n and normed by

$$||f|| = (\int_\Omega |f(\xi)|^2 d\xi)^{\frac{1}{2}}$$

The relation between the state x and the control u will be expressed by means of a linear (partial differential) equation.

2. DISTRIBUTED CONTROL, DISTRIBUTED OBSERVATION.

In this and in the following paragraph we shall consider the case where the control is distributed. We take as mathematical model of the system the following

$$\dot{x}(t) = Ax(t) + Bu(t) \qquad x(0) = 0 \tag{3}$$

where for each t the control u is an element of a Hilbert space H_2, A is the generator of a strongly continuous semigroup of linear bounded transformations $\Phi(t)$, mapping H_1 into itself, B is a bounded linear transformation mapping H_2 into $D(A)$, the domain of A. Under these assumptions (3) always admits a unique solution[3] given by

$$x(t) = \int_0^t \Phi(t-\tau)Bu(\tau)d\tau \tag{4}$$

Without loss of generality we assume that H_1 coincides with H_2. Moreover we assume that A is a linear differential operator of the form

$$A = \sum_{|j| \leq k} a_j(\xi)\, D^j$$

where the a_j's are sufficiently smooth and

$$D^j = \frac{\partial^{|j|}}{\partial \xi_1^{j_1} \ldots \partial \xi_n^{j_n}} \qquad |j| = \sum_{i=1}^{n} j_i$$

A satisfies the conditions for being the generator of a strongly continuous semi-group of linear bounded transformations[4].

We are also given the following cost functional

$$J(u;x) = \int_0^T ||x(t)-x_d(t)||^2 dt + \lambda \int_0^T ||u(t)||^2 dt \qquad (5)$$

In order for this functional to make sense we require that both u and x belong to the Hilbert space $L^2(Q)$, $Q=[0,T]x\Omega$, normed by

$$\left(\int_0^T ||f(t)||^2 dt\right)^{\frac{1}{2}}$$

Also x_d (the desired evolution) belongs to $L^2(Q)$. The class of admissible controls is characterized by a closed convex subset U of $L^2(Q)$.

The optimization problem is that of determining $u^0 \varepsilon U$, $x^0 \varepsilon L^2(Q)$ such that

$$J(u^0;x^0) = \inf_{\substack{u\varepsilon U \\ x\varepsilon L^2(Q)}} J(u;x) \quad \text{subject to (3)}$$

The ε-problem associated with it will be that of minimizing the functional

$$J_\varepsilon(u;x) = J(u;x) + \frac{1}{\varepsilon}\int_0^T ||\dot{x}(t)-Ax(t)-Bu(t)||^2 dt \qquad (6)$$

where we have chosen the functional P of (2) to be the norm over $L^2(Q)$. It remains to be specified the domain over which the minimization of J_ε is to be carried out. As far as u is concerned it will be constrained to the class of admissible controls U. As far as x is concerned it is natural to restrict it to the domain of the operator S

$$S = \frac{\partial}{\partial t} - A$$

which is defined in an obvious manner:

$$D(S) = \{f \varepsilon L^2(Q) : f \text{ is cont.diff.in } t,\ f(\cdot;0)=0,\ f(\cdot;t)\varepsilon D(A) \text{ a.e.} t\varepsilon[0,T],\ Sf\varepsilon L^2(Q)\}$$

We notice that the operator S is closable[5] and shall denote by $\bar{S}$ its closure.

We are now ready to prove the existence of solution to the ε-problem.

Theorem 1. There exist unique $u_\varepsilon \varepsilon U$ and $x_\varepsilon \varepsilon D(\bar{S})$ such that

$$J_\varepsilon(u_\varepsilon;x_\varepsilon) = J(u_\varepsilon;x_\varepsilon) + \frac{1}{\varepsilon}\int_0^T ||\bar{S}x_\varepsilon(t)-Bu_\varepsilon(t)||^2 dt = \inf_{\substack{u\varepsilon U \\ x\varepsilon D(S)}} J_\varepsilon(u;x)$$

The proof of theorem 1 can be found in S. De Julio[6] noticing that the strict convexity of J_ε accounts for the uniqueness of the solution.

We pass now to prove the convergence of u_ε and x_ε to the optimal control and trajectory, u^0, x^0, whose existence and uniqueness has been proved by A. V. Balakrishnan[3] (when the solution of (3) is taken in the strong sense) and by the author[6] (when the solution of (3) is taken in the weak sense).

Theorem 2. Let $\{\varepsilon_n\}$ be a sequence of positive numbers tending to zero. Then we have

$$\lim_{n\to\infty} u_{\varepsilon_n} = u^0 \qquad \lim_{n\to\infty} x_{\varepsilon_n} = x^0$$

Proof. It has been shown[6] that there exist subsequences $\{u_{\varepsilon_m}\}$, $\{x_{\varepsilon_m}\}$ such that

$$w\text{-}\lim_{m\to\infty} u_{\varepsilon_m} = u^0, \quad w\text{-}\lim_{m\to\infty} x_{\varepsilon_m} = x^0, \quad \lim_{m\to\infty} J_{\varepsilon_m}(u_{\varepsilon_m};x_{\varepsilon_m}) = J(u^0;x^0)$$

$$\bar{S}x^0 = u^0.$$

First of all we note that

$$\lim_{m\to\infty} \left[\int_0^T ||x_{\varepsilon_m}(t)-x_d(t)||^2dt + \lambda\int_0^T ||u_{\varepsilon_m}(t)||^2dt + \frac{1}{\varepsilon}\int_0^T ||Sx_{\varepsilon_m}(t)-Bu_{\varepsilon_m}(t)||^2dt\right]$$

$$= \int_0^T ||x^0(t)-x_d(t)||^2dt + \lambda\int_0^1 ||u^0(t)||^2dt$$

and the weak lower semi-continuity of the norm imply the strong convergence of u_{ε_m} to u^0, and of x_{ε_m} to x^0, and of $Sx_{\varepsilon_m}-Bu_{\varepsilon_m}$ to zero. Moreover from the uniqueness of the optimal control we infer that we need not take subsequences. Indeed suppose we do need, and take $\{u_{\varepsilon_m}\}$ in such a way that the subsequence left, call it $\{u_{\varepsilon_1}\}$, does not have u^0 as a cluster point. Then the application of the first part of the theorem would lead to a contradiction.

Also by contradiction we prove that x^0 is in fact the strong solution of (3) corresponding to u^0. Indeed suppose not and let $\bar{x}^0$ (it exist!) satisfy

$$S\bar{x}^0 = Bu^0$$

Then for all $\phi\in D(S^*)$ we would have

$$[x^0,S^*\phi] = [Bu^0,\phi]$$

and

$$[\bar{x}^0,S^*\phi] = [Bu^0,\phi]$$

whence

$$[x^0-\bar{x}^0, S^*\phi] = 0 \qquad \forall \phi \in D(S^*)$$

Now, with the hypotheses on A, we have that the range of S^* is dense in $L^2(Q)$. Hence the above equality implies

$$x^0 = \bar{x}^0$$

It is probably worth noting that having taken the initial condition in (3) to be zero is not a restriction at all, since we can always reduce ourselves to this case by means of the following artifice. Let the initial condition be

$$x(0) = x_0 \in D(A^2)$$

and define a new state variable w

$$w(\xi;t) = x(\xi;t) - x_0(\xi)$$

Differentiating with respect to t and taking (3) into account we get

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

or

$$\dot{w} = A\,w + B\,u + A\,x_0 \qquad w(0) = 0$$

Being Ax_0 in the domain of A, the addition of the new forcing term does not raise any difficulty.

3. DISTRIBUTED CONTROL, FINAL OBSERVATION.

Let (3) be the system equation and let the same hypotheses of the previous section hold. Instead of (5) we here considere the following cost functional

$$J(u;x) = ||x(T)-x_d||^2 + \lambda \int_0^T ||u(t)||^2 dt \qquad (7)$$

where now x_d belongs to $L^2(\Omega)$.

As in (6) we define the functional J_ε as follows

$$J_\varepsilon(u;x) = J(u;x) + \frac{1}{\varepsilon} \int_0^T ||\dot{x}(t)-Ax(t)-Bu(t)||^2 dt$$

With the new definition of the cost functional theorems 1 and 2 still apply verbatim, but the proofs need some more work. We shall here give in full the proof of theorem 1.

In order not to make the proof to cumbersome we first prove the following lemma.

Lemma. Given the differential equation

$$\dot{x}(t) = Ax(t) + Bu(t) + z(t) \qquad x(0)=0 \tag{8}$$

where A and B satisfy the conditions stated above, then its weak solution x is such that the mapping $(u,z)\to x$ $(L^2(Q)xL^2(Q)\to L^2(Q))$ as well as the mapping $(u,z)\to x(T)$ $(L^2(Q)xL^2(Q)\to L^2(\Omega))$ are both continuous.

Proof. Equation (8) does not necessarily admit a strong solution since $z(t)$ is not assumed to lie in the domain of A ; but it will always have a weak solution satisfying

$$\overline{S}x(t) = Bu(t) + z(t) \qquad x(0)=0$$

given by

$$x(t) = \int_0^t \Phi(t-\tau)Bu(\tau)d\tau + \int_0^t \Phi(t-\tau)z(\tau)d\tau \tag{9}$$

In particular we have

$$x(T) = \int_0^T \Phi(T-\tau)Bu(\tau)d\tau + \int_0^T \Phi(T-\tau)z(\tau)d\tau \tag{10}$$

and both the mappings $(u,z)\to x$ and $(u,z)\to x(T)$ defined by (9) and (10) are continuous.

Proof of theorem 1. Let $\{u_n,x_n\}$ be a sequence minimizing J_ε :

$$J_\varepsilon(u_n;x_n) \to j_\varepsilon = \inf_{\substack{u\in U \\ x\in D(S)}} J_\varepsilon(u;x)$$

and let

$$z_n = Sx_n - Bu_n$$

There obviously exist constants C_1 , C_2 , C_3 such that

$$\int_0^T ||u_n(t)||^2dt \leq C_1$$

$$\int_0^T ||z_n(t)||^2dt \leq C_2$$

$$\int_0^T ||Sx_n(t)||^2dt \leq C_3$$

Therefore, there exist subsequences $\{u_m\}$, $\{x_m\}$, $\{z_m\}$ such that

$$\underset{m\to\infty}{\text{w-lim}}\ u_m = u_\varepsilon \in U$$

$$\underset{m\to\infty}{\text{w-lim}}\ Sx_m = y_\varepsilon \in L^2(Q)$$

$$\underset{m\to\infty}{\text{w-lim}}\ z_m = z_\varepsilon \in L^2(Q)$$

Applying the previous lemma we also have

$$\underset{m\to\infty}{\text{w-lim}}\ x_m = x_\varepsilon$$

$$\underset{m\to\infty}{\text{w-lim}}\ x_m(T) = x_\varepsilon(T)$$

where x_ε satisfies

$$\bar{S}x_\varepsilon(t) = Bu_\varepsilon(t) + z_\varepsilon(t)$$

It can also be shown[6] that

$$y_\varepsilon = \bar{S}x_\varepsilon$$

Applying now the weak lower semi-continuity of J_ε, we get

$$j_\varepsilon = \lim_{m\to\infty} J_\varepsilon(u_m;x_m) \geq J_\varepsilon(u_\varepsilon;x_\varepsilon)$$

for which only equality can hold.

In much the same way we can extend the proof of theorem 2 to the present case.

4. BOUNDARY CONTROL.

We now consider the case where the control u is on the boundary. Namely we consider systems described by an equation of the type

$$\begin{aligned} \dot{x}(t) &= Ax(t) \qquad x(0)=0 \\ x|_\Sigma &= u \end{aligned} \tag{11}$$

where $\Sigma=\Gamma x[0,T]$, Γ being the boundary of Ω. So far we have not made any assumption on Γ. Here we assume that Γ is bounded and "smooth".

The constraint set U for the control is now a closed convex subset of $L^2(\Sigma)$.

We denote by $V(Q)$ the space of functions $f \in L^2(Q)$ normed by

$$||f|| = \left(\int_0^T [||f(t)||^2 + ||Sf(t)||^2]dt\right)^{\frac{1}{2}}$$

It is known[7] that V is a Hilbert space. We assume that V is algebraically and topologically contained in $H^1(Q)$ (the Sobolev space of order 1).

Let γ be the trace operator defined by

$$\gamma f = f|_\Sigma$$

Then, with the above assumptions, γ is continuous as a mapping from $V(Q)$ into $L^2(\Sigma)$[8].

Referring to the case of distributed observation, we take as cost functional the following

$$J(u;x) = \int_0^T ||x(t)-x_d(t)||^2 dt + \lambda \int_0^T ||u(t)||_1^2 dt$$

where $||\cdot||_1$ denotes the norm of $L^2(\Sigma)$. The functional J_ε associated with the ε-problem will then be

$$J_\varepsilon(u;x) = J(u;x) + \frac{1}{\varepsilon}\int_0^T ||\dot{x}(t)-Ax(t)||^2 dt$$

We now pass to prove theorems 3 and 4 which are the counterparts of theorems 1 and 2.

Theorem 3. There exist unique $u_\varepsilon \in L^2(\Sigma)$ and $x_\varepsilon \in D(\overline{S})$ such that

$$J_\varepsilon(u_\varepsilon;x_\varepsilon) = J(u_\varepsilon;x_\varepsilon) + \frac{1}{\varepsilon}\int_0^T ||\overline{S}x_\varepsilon(t)||^2 dt = \inf_{\substack{u\in U \\ x\in D(S)}} J_\varepsilon(u;x) \text{ subject to } x|_\Sigma = u$$

Proof. Let $\{u_n,x_n\} \in U \times D(S)$ $(u_n = x_n|_\Sigma)$ be a minimizing sequence, i.e.

$$J_\varepsilon(u_n;x_n) \to j_\varepsilon = \inf_{\substack{u\in U \\ x\in D(S)}} J_\varepsilon(u;x) \text{ subject to } x|_\Sigma = u$$

Then there exist constants C_1, C_2, C_3 such that

$$\int_0^T ||u_n(t)||^2 dt \leq C_1$$

$$\int_0^T ||x_n(t)||^2 dt \leq C_2$$

$$\int_0^T ||Sx_n(t)||^2 dt \leq C_3$$

Therefore there exist subsequences such that

$$\underset{m\to\infty}{w\text{-}\lim}\ u_m = u_\varepsilon \in U$$

$$\underset{m\to\infty}{\text{w-lim}}\, x_m = x_\varepsilon \;\varepsilon L^2(Q)$$

$$\underset{m\to\infty}{\text{w-lim}}\, Sx_m = y_\varepsilon \;\varepsilon L^2(Q)$$

It can be shown[6] that

$$x_\varepsilon \varepsilon D(\bar{S}) \quad \text{and} \quad y_\varepsilon = \bar{S}x_\varepsilon$$

Now, since x_n converges weakly in V, by the continuity of the mapping γ we also have

$$x_\varepsilon|_\Sigma = u_\varepsilon$$

Exploiting the weak lower semi-continuity of J_ε we get

$$j_\varepsilon = \lim_{m\to\infty} J_\varepsilon(u_m;x_m) \geq J_\varepsilon(u_\varepsilon;x_\varepsilon) \qquad x_\varepsilon|_\Sigma = u_\varepsilon$$

for which only equality can hold.

Theorem 4. Let $\{\varepsilon_n\}$ be a sequence of positive numbers tending to zero. Then we have

$$\lim_{n\to\infty} u_{\varepsilon_n} = u^0 \qquad \lim_{n\to\infty} x_{\varepsilon_n} = x^0$$

where u^0 and x^0 are the optimal control and trajectory.

Proof. Taking again into account the continuity of the mapping γ and the proof of theorem 2, this theorem can be easily demonstrated.

We note that the remark on the vanishing of the initial condition made at the end of section 2 applies also to this case. Moreover the presence of a forcing term in (11) would not affect the proofs of theorems 3 and 4 either.

Finally we observe that also the case of final observation could be easily carried out.

5. A COMPUTATIONAL EXAMPLE.

The computational technique proposed has been applied to a particular example[9] of distributed parameter system. We have considered the problem of determining the minimal norm control for the system

$$\frac{\partial x}{\partial t} = \alpha \frac{\partial^2 x}{\partial \xi^2} + u(\xi;t) \qquad 0 \leq \xi \leq 1 \qquad 0 \leq t \leq 5$$

$$x(0;t) = x(1;t) = 0 \tag{12}$$

$$x(\xi;0) = x_o(\xi) \qquad x(\xi;5) = x_1(\xi)$$

The initial and final state, x_o and x_1 are given in fig1. The numerical value of α is $\alpha = .0033$.

The ε-problem associated with this case is that of minimizing the functional J_ε given by

$$J_\varepsilon(u;x) = \int_0^5 dt \int_0^1 u^2(\xi;t)d\xi + \frac{1}{\varepsilon}\int_0^5 dt \int_0^1 (\dot{x}(\xi;t) - \frac{\partial^2 x}{\partial \xi^2}(\xi;t) - u(\xi;t))^2 d\xi \tag{13}$$

This problem slightly differs from the one presented in paragraph 2 for the absence of the term involving the desired state. But it can be easily shown that, because of the continuity of the mapping defined by (4), theorems 1 and 2 still hold.

In order to solve the problem on a digital computer we have discretized both space and time variable. Let $\Delta\xi$ and Δt be the discretization intervals, with

$$N\Delta\xi = 1 \qquad M\Delta t = 5$$

and denote

$$u_{ij} = u(j\Delta\xi;i\Delta t)$$
$$x_{ij} = x(j\Delta\xi;i\Delta t) \qquad i=0,\ldots,M \; ; \quad j=0,\ldots,N$$

The discrete version of (13) will then be

$$J_\varepsilon = \Delta\xi\Delta t \sum_{i=0}^{M-1}\sum_{j=1}^{N-1} u_{ij}^2 + \frac{\Delta\xi\Delta t}{\varepsilon}\sum_{i=0}^{M-1}\left(\frac{x_{i+1,j}-x_{ij}}{\Delta t} - \alpha\,\frac{x_{i,j-1}-2x_{ij}+x_{i,j+1}}{\Delta\xi^2} - u_{ij}\right)^2 \tag{14}$$

where we have used the forward difference and the second central difference approximation for the first and second order derivatives respectively.

The minimizing values of u_{ij} and x_{ij} are determined by imposing that the gradient of J_ε with respect to u and x vanishes. Explicitly we have

$$\frac{1}{\Delta\xi\Delta t}\frac{\partial J_\varepsilon}{\partial u_{ij}} = 2u_{ij} - \frac{2}{\varepsilon}\left(\frac{x_{i+1,j}-x_{ij}}{\Delta t} - \alpha\,\frac{x_{i,j-1}-2x_{ij}+x_{i,j+1}}{\Delta\xi^2} - u_{ij}\right) = 0 \tag{15}$$

$$i=0,\ldots,M-1 \; , \quad j=1,\ldots,N-1$$

$$\frac{\varepsilon}{2\Delta\xi\Delta t}\frac{\partial J_\varepsilon}{\partial x_{ij}} = \frac{1}{\Delta t}\left(\frac{x_{ij}-x_{i-1,j}}{\Delta t} - \alpha\,\frac{x_{i-1,j-1}-2x_{i-1,j}+x_{i-1,j+1}}{\Delta\xi^2} - u_{i-1,j}\right)$$

$$+\left(\frac{2\alpha}{\Delta\xi^2}-\frac{1}{\Delta t}\right)\left(\frac{x_{i+1,j}-x_{ij}}{\Delta t} - \alpha\,\frac{x_{i,j-1}-2x_{ij}+x_{i,j+1}}{\Delta\xi^2} - u_{ij}\right) \qquad (16)$$

$$-\frac{\alpha}{\Delta\xi^2}\left(\frac{x_{i+1,j+1}-x_{i,j+1}}{\Delta t} - \alpha\,\frac{x_{ij}-2x_{i,j+1}+x_{i,j+2}}{\Delta\xi^2} - u_{i,j+1}\right)$$

$$-\frac{\alpha}{\Delta\xi^2}\left(\frac{x_{i+1,j-1}-x_{i,j-1}}{\Delta t} - \alpha\,\frac{x_{i,j-2}-2x_{i,j-1}+x_{ij}}{\Delta\xi^2} - u_{i,j-1}\right) = 0$$

$$i=1,\ldots,N-1\ ,\quad j=2,\ldots,N-2$$

$$\frac{\varepsilon}{2\Delta\xi\Delta t}\frac{\partial J_\varepsilon}{\partial x_{i1}} = \frac{1}{\Delta t}\left(\frac{x_{i1}-x_{i-1,1}}{\Delta t} - \alpha\,\frac{x_{i-1,0}-2x_{i-1,1}+x_{i-1,2}}{\Delta\xi^2} - u_{i-1,1}\right)$$

$$+\left(\frac{2\alpha}{\Delta\xi^2}-\frac{1}{\Delta t}\right)\left(\frac{x_{i+1,1}-x_{i1}}{\Delta t} - \alpha\,\frac{x_{i0}-2x_{i1}+x_{i2}}{\Delta\xi^2} - u_{i1}\right) \qquad (16')$$

$$-\frac{\alpha}{\Delta\xi^2}\left(\frac{x_{i+1,2}-x_{i2}}{\Delta t} - \alpha\,\frac{x_{i1}-2x_{i2}+x_{i3}}{\Delta\xi^2} - u_{i2}\right) = 0$$

$$i=1,\ldots,N-1$$

$$\frac{\varepsilon}{2\Delta\xi\Delta t}\frac{\partial J_\varepsilon}{\partial x_{i,N-1}} = \frac{1}{\Delta t}\left(\frac{x_{i,N-1}-x_{i-1,N-1}}{\Delta t} - \alpha\,\frac{x_{i-1,N-2}-2x_{i-1,N-1}-x_{i-1,N}}{\Delta\xi^2} - u_{i-1,N-1}\right)$$

$$+\left(\frac{2\alpha}{\Delta\xi^2}-\frac{1}{\Delta t}\right)\left(\frac{x_{i+1,N-1}-x_{i,N-1}}{\Delta t} - \alpha\,\frac{x_{i,N-2}-2x_{i,N-1}+x_{iN}}{\Delta\xi^2} - u_{i,N-1}\right) \qquad (16'')$$

$$-\frac{\alpha}{\Delta\xi^2}\left(\frac{x_{i+1,N-2}-x_{i,N-2}}{\Delta t} - \alpha\,\frac{x_{i,N-3}-2x_{i,N-2}+x_{i,N-1}}{\Delta\xi^2} - u_{i,N-2}\right) = 0$$

$$i=1,\ldots,N-1$$

Now, since (15) can easily be solved for u_{ij}:

$$u_{ij} = \frac{1}{1+\varepsilon}\left(\frac{x_{i+1,j}-x_{ij}}{\Delta t} - \alpha\,\frac{x_{i,j-1}-2x_{ij}+x_{i,j+1}}{\Delta\xi^2}\right) \qquad (17)$$

whereas (16), (16'), (16") are of less easy solution, we have chosen an iterative procedure according to the following steps:

1. x is initially guessed.
2. u is computed by means of (17).
3. The gradient technique is applied to determine the minimizing value of x for the computed u .

Step 2 and 3 are repeated alternately until the global minimum is reached.

In fig.2 the flow chart for the program implemented on a digital computer is shown. This flow chart is self-explanatory.

In the final program that was run on the computer, we have actually decided to abolish cycle number 2, due to the fact that, through computational experience, we have noticed that nothing is gained by computing the minimizing value of x for each u, but it sufficed to determine the minimizing value of x in the direction of the gradient. In fact the computing time was considerably reduced in this way.

Figures 3 and 4 show the results of the numerical solution of the ε-problem with $\varepsilon=.01$. Also the exact solution x corresponding to the u_ε plotted in fig.3 was computed. The relative diagram has not been plotted because the difference from the x_ε shown in fig.4 could not be appreciated. In fact the maximum difference between x and x_ε was less than 2%.

6. CONCLUSIONS.

The computational technique that we have proposed seems to be very powerful. In this work we have shown its applicability to a particular class of linear systems described by partial differential equations, with a particular class of cost functionals, namely quadratic criteria.

If we are willing to accept weak convergence of the solution of the ε-problem to the optimal solution, instead of strong convergence, a more general type of cost functionals can be handled by this technique[6]. Moreover the method seems to be powerful also from a theoretical standpoint. Balakrishnan[1] has shown that an easy proof of the Maximum Principle can be given and the author[6] has actually proved the existence of optimal controls for a broad class of infinite dimensional linear systems.

If we turn to non-linear systems we see that in order for the technique to apply we have to considerably restrict the class of systems[1]. We hope that further research will enable us to relax such restrictions.

This research was partially supported by AFOSR under Grant 68-1408 Applied Mathematics Division, USAF and by the Consiglio Nazionale delle Ricerche.

REFERENCES

[1] A.V.BALAKRISHNAN, "On a New Computing Method in Optimal Control," Uni-

versity of California at Los Angeles, Department of Engineering, Report No.67-61, December, 1967.

[2] S.DE JULIO,"Computation of Optimal Controls for Infinite Dimensional Systems," Paper presented at the Second Annual Princeton Conference on Information Sciences and Systems, March 25-26, 1968, Princeton, New Jersey.

[3] A.V.BALAKRISHNAN, "Optimal Control Problems in Banach Spaces", J. SIAM Control, Ser.A, Vol.3, No.1, 1965.

[4] N.DUNFORD and J.T.SCHWARTZ, Linear Operators, Part I, New York. Interscience Publisher Inc., 1958.

[5] K.YOSIDA, Functional Analysis, Springer-Verlag, 1965.

[6] S.DE JULIO, "On the Optimization of Infinite Dimensional Linear Systems," Paper presented at the 2nd International Conference on Computing Methods in Optimization Problems, September 9-13, 1968, San Remo, Italy.

[7] J.L.LIONS, Equations Differentielles Operationnelles, Springer-Verlag, 1961.

[8] J.DENY and J.L.LIONS, "Les Espaces du Type de Beppo-Levi," Annales Fourier, Vol.V, 1955.

[9] D.A.WISMER, Jr.,"Optimal Control of Distributed Parameter Systems Using Multilevel Techniques," University of California at Los Angeles, Department of Engineering, Report No.66-55, November, 1966.

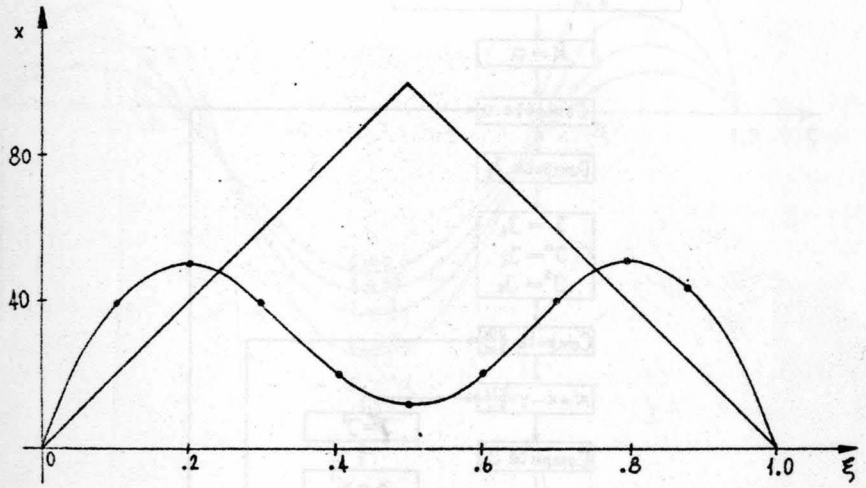


Fig. 1.

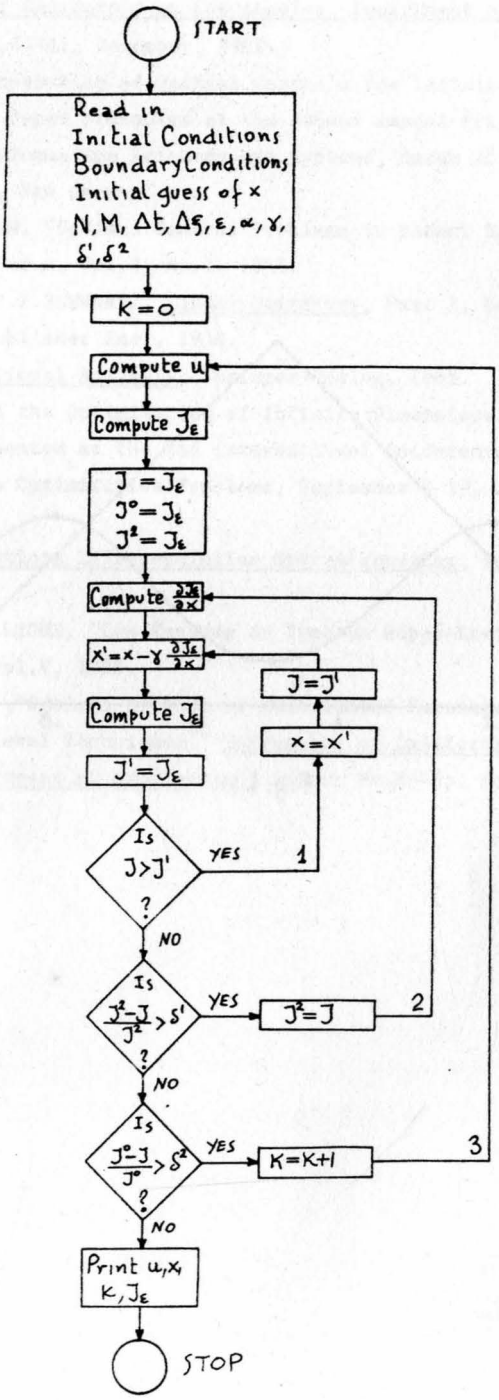

START
Read in
Initial Conditions
Boundary Conditions
Initial guess of x
N, M, Δt, Δξ, ε, α, γ,
δ', δ²
K=0
Compute u
Compute Jε
J = Jε
J° = Jε
J² = Jε
Compute ∂Jε/∂x
x' = x − γ ∂Jε/∂x
J = J'
Compute Jε
x = x'
J' = Jε
Is J > J' ?
YES
1
NO
Is (J² − J)/J² > δ' ?
YES
J² = J
2
NO
Is (J° − J)/J° > δ² ?
YES
K = K+1
3
NO
Print u, x, K, Jε
STOP

Fig. 2.

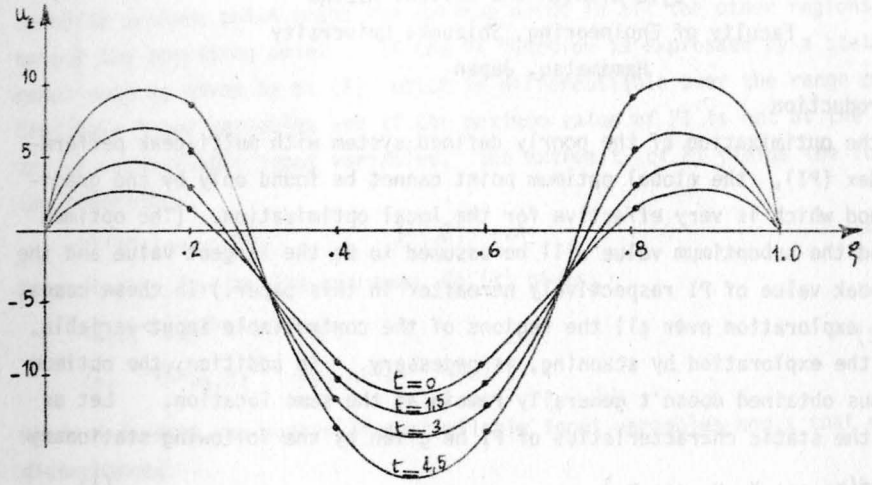

u_ξ
10
5
-5
-10
.2
.4
.6
.8
1.0
ξ
t=0
t=1.5
t=3
t=4.5

Fig. 3

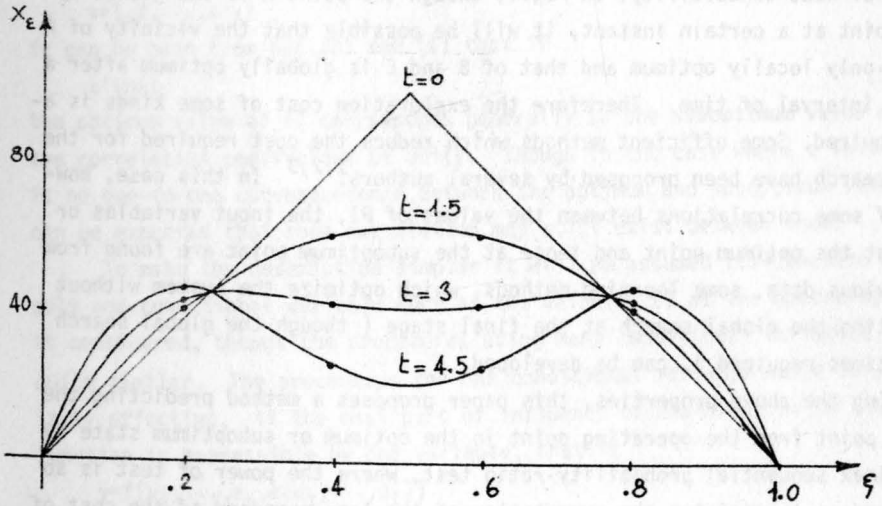

x_ξ
80
40
t=0
t=1.5
t=3
t=4.5
.2
.4
.6
.8
1.0
ξ

Fig. 4

A STATISTICAL APPROACH TO THE OPTIMIZATION OF THE CONTROL SYSTEMS WITH MULTI-PEAK PERFORMANCE INDEX

Fujio NISHIDA and Seiichi MIZUNO
Faculty of Engineering, Shizuoka University
Hamamatsu, Japan

1. Introduction

In the optimization of the poorly defined system with multi-peak performance index (PI), the global optimum point cannot be found only by the gradient method which is very effective for the local optimization. (The optimum value and the suboptimum value will be assumed to be the largest value and the second peak value of PI respectively hereafter in this paper.) In these cases, a global exploration over all the regions of the controllable input variable, such as the exploration by scanning, is necessary. In addition, the optimum point thus obtained doesn't generally remain at the same location. Let us suppose the static characteristics of PI be given by the following stationary model: $y=f(x_1,\cdots,x_n,u_1,\cdots,u_\ell)$ (1)
where $x_1,\cdots,x_n$ are controllable input variables, $u_1,\cdots,u_\ell$ are directly unmeasurable disturbances such as the variation of percentage impurity ratio or that of heat conductivity. In Fig.1, though the point A is the global optimum point at a certain instant, it will be possible that the vicinity of A becomes only locally optimum and that of B and C is globally optimum after a certain interval of time. Therefore the exploration cost of some kinds is again required. Some efficient methods which reduce the cost required for the global search have been proposed by several authors.[1,2,3] In this case, however, if some correlations between the values of PI, the input variables or others at the optimum point and those at the suboptimum point are found from the previous data, some learning methods, which optimize the system without prosecuting the global search at the final stage (though the global search is sometimes required), can be developed.

Using the above properties, this paper proposes a method predicting the optimum point from the operating point in the optimum or suboptimum state by the minimax sequential probability-ratio test, where the power of test is so determined as to minimize the expectation of the loss composed of the cost of the global exploration and the risk due to the incorrect decision.

According to the above mentioned procedures, simulation was performed with a computer. The results of the simulation coincide fairly well with the analytical result, showing that the expectation of the loss can be much reduced if any correlation is available.

2. The possibility of the existence of correlation

Let us consider the possibility of the existence of the correlation between the values of PI or the input variable at the opposite optimum point and those at the operating point which is assumed to be locally optimum. The opposite optimum point means the optimum point in all the other regions except the operating point. If the PI function is expressed by a stationary model such as given by Eq.(1) which is differentiable over the range of controllable input variables and if the maximum value of PI is not at the boundary of controllable input variables, the extremity of PI yields the following condition.

$$\partial f/\partial x_j|_{x=x_{e,j}}=0 \qquad j=1,2,\cdots,n \tag{2}$$

Accordingly, for the ith extremum, Eq.(2) gives;

$$\begin{aligned} x_{e_i,j}&=\phi_{i,j}(u_1,\cdots,u_\ell) \\ y_{e_i}&=f(x_{e_i,1},\cdots,x_{e_i,n},u_1,\cdots,u_\ell) \end{aligned} \tag{3}$$

where n denotes the number of controllable input variables and ℓ that of disturbances.

Furthermore, assuming that there exist some measurable quantities v_s that are independent of the value of PI as well as of each other, namely:

$$v_s=\Phi_s(x_1,\cdots,x_n,u_1,\cdots,u_\ell) \tag{4}$$
$$s=1,2,\cdots,\ell'-1$$

it can be seen from Eqs.(3) and (4) that if

$$\ell'+n\geq\ell \tag{5}$$

the optimum value of PI corresponds generally to the suboptimum value with the correlation coefficient of unity. Though in the case where $\ell'+n<\ell$ there is no one-to-one correspondence between the optimum and suboptimum value, it can be expected that some correlation may still exist between them.

To make the description simpler it will be assumed till section 8 that only one conditional variable such as the value of PI at the operating point is considered, though the procedures using many conditional variables are quite similar. The procedures for one conditional variable would be particularly effective, if the most part of influence of disturbances to the PI function is expressible by one variable, that is

$$y=f(x_1,\cdots,x_n,\Phi(u_1,\cdots,u_\ell)) \tag{6}$$

3. The global exploration

In order to obtain sufficient data refering to the pairs of the optimum and the suboptimum point, a certain efficient global exploration is required to be performed. At the end of each global exploration in a certain state of

the plant, the PI value and the input variable at the optimum and those at the suboptimum point are recorded in relation to the classified value of PI at the suboptimum and that at the optimum point respectively, and then the value of Z is calculated from Eq.(27) which will be shown later. Thereafter the system operates at the optimum point. Here the classification is necessary to restrict the number of classes of the conditional probability density function to an appropriate magnitude, and this classified value is used as the conditional value for the probability. After a certain interval of time T_s (sampling interval), or in the case where any change of the plant state is detected from the variations of some quantities at the operating point, the new global exploration is again initiated. In the next paragraph some methods of the exploration will be considered.

3.1 Exploration by a scanning

This method, in principle, is the one to scan over all the range of the controllable input variable, and it is necessary for the state of the plant to remain almost unchanged until the end of the scanning. Therefore, in order to increase the exploration speed, the coorperation with some other efficient procedures such as a polynomial approximation of the PI function will be preferable.

3.2 The exploration by an estimation

The exploration by the above method requires much time if there are numbers of the controllable input variables. Here, an exploration method by an estimation will be discussed. Let us devide all the region of the controllable input variables into the subdomains of a proper size. The system explores the PI values of the unknown subdomain by means of the estimation or the measurement based on the optimum point that has been already known.

3.2.1 The estimation of PI value in the unknown subdomain near the known one

We shall assume that after obtaining the measured PI value y_i of a certain subdomain in a state of the plant, the absolute difference of PI $|\Delta y| = |y_i - y_j|$ and the distance d are calculated from the measured values y_j and x_j in the known subdomain in the neighbourhood and then the average value of $|\Delta y|$ as well as the dispersion versus the distance d is recorded. In order to estimate the PI value in the unknown subdomain, the following relation is examined by using the PI value y_o in the nearest known subdomain:

$$\begin{aligned} &P_r(y_u \leq y'_m) > \gamma_1 \\ &y'_m = y_m - \delta \end{aligned} \tag{7}$$

where y_u corresponds to the upper bound of the unknown PI value, y_m is the optimum value that has already been found in the present state, δ is a positive constant, and the probability density function P_r is assumed to

be normal. The value of the threshold γ_1 is desired to be determined so as to minimize the loss due to the estimation, and the loss will be approximately minimized if δ is small and if the value of γ_1 is taken so that the expectation of the loss due to measuring in spite of $y<y_m$ equals that due to the neglection of measuring in spite of $y \geq y_m$; that is

$$(1-\gamma_1)(y_{max}-y_m')=\gamma_1 C_m$$

$$\gamma_1=\frac{y_{max}-y_m'}{y_{max}-y_m'+C_m} \tag{8}$$

where y_{max} is the maximum value of PI that has been known throughout all the states and C_m the cost of the measurement. Now, if the relation (7) holds, the exploration for the other unknown subdomains is initiated without measurements. Otherwise the following estimation is to be performed.

3.2.2 Estimation of the range governed by a PI hill

As can be seen in Fig.2, the PI value of the subdomain C and D located near the peak A or B, respectively, will be comparable with that of the subdomain A or B, but the former will, in most cases, be less than the latter because the local optimization is performed near the peak, and hence no measurement in the unknown subdomain near the peak will be preferable. Based on the above consideration, it is assumed that for every time when any non trivial peaks such as the optimum or suboptimum ones are recorded, the average distance d between a nearest pair of peaks as well as the dispersion is calculated and that d/2 is the mean value of the radius of the region governed by each peak.

Now, let's suppose an unknown subdomain separated by r from a certain known peak having the unknown radius x of the governing region. If the relation

$$P_r(r<x)>\gamma_2 \tag{9}$$

is satisfied, the exploration of the other unknown subdomain is initiated without measurement. This will here be called the beginning procedure. Otherwise, the system measure the unknown PI value y and if $y<<y_m$ the system returns to the above beginning procedure. On the other hand, if $y \simeq y_m$ or $y>y_m$, the corresponding local optimum point is searched by the local optimization method and the data as well as the value of Z will be recorded, and thereafter the system returns to the beginning procedure. The threshold value of γ_2 can be determined in a similar way as γ_1 in Eq.(8).

4. The application procedures of the sequential probability ratio test

Now, let's show the procedures of the statistical decision. It is here assumed that the opposite optimum value of PI y has a conditional probabili-

ty density distribution function $f(y,\theta|y^*)$, where y^* denotes the classified PI value of the operating point which is assumed to be locally optimum and θ is the mean value of y. Adopting the null hypothesis

$$\theta \leq y' \qquad (y' \geq y^*) \tag{10}$$

this problem is reduced to be that of the hypothesis test. The test of this composite hypothesis, as is well known, becomes to that of the simple null hypothesis

$$\theta = \theta_o = y' - d \tag{10'}$$

d; a proper positive constant

for the alternative

$$\theta = \theta_1 = y' + d \tag{11}$$

In this problem, it will be suitable to take y' and d as follows

$$\begin{aligned} y' &= y^* + C_2 \\ d &= C_2 \end{aligned} \tag{11'}$$

where C_2 is the local optimization cost. The above procedures would examine their relative magnitude of likelihood rather than the validity of the proposition $\theta=\theta_o$ or $\theta=\theta_1$.

Now, let us assume that a form of the risk due to the incorrect decision $\theta=\theta_o$ where $\theta=\theta_1$ is true as is shown in the following equation:

$$w_{10} = m_1 \{ \int_{y^*}^{\infty} y f(y,\theta_1|y^*)dy - y^* \} \tag{12}$$

where m_1 is the expected number that this incorrect decision is adopted without correction, and is proportional to the duration T_a. (See Fig.3.) Similarly, let us assume that the following simplified equation may denote the risk due to the incorrect decision $\theta=\theta_1$ when $\theta=\theta_o$ is true, though it may be more complicated in the actual plant:

$$w_{01} = m_2 [C_2 + k\{y^* - \int_{-\infty}^{y^*} y f(y,\theta_o|y^*)dy\}] \tag{13}$$

$$0<k<1$$

where C_2 is the locally exploring cost near the opposite optimum point, which is generally less than the global exploration cost C_1, and m_2 is the expected number of this incorrect decision-applying, where the incorrect decision can be corrected by continuing the measurement and the test, accordingly, m_2 will generally be less than m_1.

Now, the procedures of testing the simple null hypothesis for the alternative is treated by the following algorithm. If the apriori probability distribution $\lambda=(p,1-p)$ for $\theta=\theta_i$ $(i=0,1)$ belongs to the set where the loss is less than that due to any decision through one measurement at least, the decision $\theta=\theta_i$ $(i=0,1)$ is immediately made without measurement. Otherwise, using the measured values $y=y_1,\cdots,y_k$ and the properly chosen constants A, B

$(A>B>0)$,

if $P_{1k}/P_{0k} \geq A$, $\theta=\theta_1$ is decided; decision $d_e=d_1$

if $P_{1k}/P_{0k} \leq B$, $\theta=\theta_0$ is decided; decision $d_e=d_0$

if $A>P_{1k}/P_{0k}>B$, measurement is to be continued in order to obtain more measured data; decision $d_e=d_2$ (14)

where

$$P_{1k}=f(y_1,\theta_1|y^*)\cdots f(y_k,\theta_1|y^*)$$
$$P_{0k}=f(y_1,\theta_0|y^*)\cdots f(y_k,\theta_0|y^*) \quad (15)$$

Let the expectation of loss due to the decision $\theta=\theta_j$ where $\theta=\theta_i$ be R_i $(j\neq i, i=0,1)$, the probability of the incorrect decision $\theta=\theta_j$ be α and β, the average number of global exploration necessary to the decision for $\theta=\theta_i$ be n_i, and the global exploration cost be C_1. Then the expectation of the loss resulted from the decision due to the sequential probability ratio test is

$$R=pR_0+(1-p)R_1 \quad (16)$$

where

$$R_0=\alpha W_{01}+C_1 n_0 \quad (17)$$
$$R_1=\beta W_{10}+C_1 n_1 \quad (18)$$

The sequential probability ratio test is, as is well known, the sequential test that minimizes R, and the decision procedures are called the Bays' solution for the apriori probability distribution λ. Now, if the value of p is known in Eq.(16), the value of A and B minimizing R can be determined. In most cases, however, the value of p is unknown. Accordingly the application of the minimax strategy, in which the maximum of R for the value of p is minimized, is considered. In order to obtain the minimax decision rule, we equate R_0 and R_1 as is known in the decision theory. On the basis of the above relation, the optimum threshold values A and B that minimize $R_0(=R_1)$ can be determined. Here α, β, n_0 and n_1 are, as is well known, approximately given as follows;

$$\alpha \simeq (1-B)/(A-B)$$
$$\beta \simeq B(A-1)/(A-B) \quad (19)$$

or

$$A \simeq (1-\beta)/\alpha$$
$$B \simeq \beta/(1-\alpha) \quad (20)$$
$$n_0 \simeq \{\alpha \log A+(1-\alpha)\log B\}/E_{\theta_0}$$
$$n_1 \simeq \{(1-\beta)\log A+\beta \log B\}/E_{\theta_1} \quad (21)$$
$$Z=\log\{f(y,\theta_1|y^*)/f(y,\theta_0|y^*)\} \quad (22)$$

where E_{θ_i} is the expectation of Z for $\theta=\theta_i$ $(i=0,1)$. Thus, if the sequential probability ratio reaches the threshold value A or B, the decision whether

to accept or to reject the hypothesis of $\theta=\theta_o$ is made and then the system operates at the optimum point.

In the next sampling duration T_s, the value of PI at the operating point is measured and the system examines whether or not a decision regarding the class has already been made. If the decision belongs to the accepting, in other words, if the operating point is probably optimum, the system is kept near the present operating point with the local optimization. On the contrary, if the decision belongs to the rejecting, the operating point would jump to the probable optimum point whose coordinate will be given by the average value corresponding to the input variable, and then the optimum point will be searched locally. On the other hand, when a decision is not made, the system explores globally as mentioned above. These decision-applying procedures are carried out during the period T_a.

Now, if the recalculation of Z is performed after the decision $d_e=d_1$ when the operating point moves to the opposite optimum subdomain, it can be examined whether or not the same decision is made. Accordingly, at the recalculation cost of Z, the risk due to the decision $d_e=d_1$ can be reduced. Furthermore, by the estimation of the stationary characteristic variation detected from the recalculation results of Z for the decision $d_e=d_1$, the system can again enter the global exploration period, if desired. Therefore the predetermined decision-applying period T_a can be increased if the recalculation of Z is performed.

5. The minimax numerical solution

If $f(y,\theta|y^*)$ is given by the following normal probability density function with the mean value θ and the dispersion σ^2:

$$f(y,\theta)=(2\pi\sigma^2)^{-\frac{1}{2}}\,\varepsilon^{-(y-\theta)^2/(2\sigma^2)} \tag{23}$$

then, from Eq.(22)

$$Z=\{2(\theta_1-\theta_0)y+\theta_0{}^2-\theta_1{}^2\}/(2\sigma^2) \tag{24}$$

Hence, the expectation E_{θ_0} and E_{θ_1} for $\theta=\theta_0$ and $\theta=\theta_1$ respectively are expressed as follows

$$E_{\theta_0}=-E_{\theta_1}=-(\theta_0-\theta_1)^2/(2\sigma^2) \tag{25}$$

Putting

$$d=\theta_1-y'=y'-\theta_0\,. \tag{26}$$

Z becomes

$$Z=2d(y-y')/\sigma^2 \tag{24'}$$

Defining

$$S=\sum_{i=1}^{m} Z_i=(2d/\sigma^2)\sum_{i=1}^{m}(y_i-y') \tag{27'}$$

the conditions of Eq.(14) are replaced by

if $S \geq \log A$, $\theta=\theta_1$ is decided; decision $d_e=d_1$

if $S \leq \log B$, $\theta=\theta_0$ is decided; decision $d_e=d_0$ (27)

if $\log A > S > \log B$, measurement is to be coutinued; decision $d_e=d_2$

Normalizing R_1, W_{01} and W_{10} as follows

$$r=R_1/(C_1/E_{\theta_1})$$
$$w_o=W_{01}/(C_1/E_{\theta_1}) \quad (28)$$
$$u=W_{10}/W_{01}$$

the minimax condition $R_0=R_1$, from Eqs.(17) and (18), yields

$$B(A-1)uw_0-(1-B)w_0+(1+A)(1-B)\log A+(1+B)(A-1)\log B=0 \quad (29)$$

and

$$r=\{(1-B)w_0-(1-B)\log A-(A-1)\log B\}/(A-B) \quad (30)$$

Equation (29) is solvable in the case of u=1 and the solution is

$$B=1/A \quad (31)$$

hence, from Eq.(19)

$$\beta=\alpha \quad (32)$$

and from Eq.(30), r is

$$r=\{w_0+(A-1)\log A\}/(A+1) \quad (33)$$

Figure 4a shows the minimum value of r and the corresponding value of A versus w_0 in the case of u=1. From Fig.4a it is evident that, the larger w_0 ($\propto W/C_1$) is, the larger A (=1/B) becomes. In other words, the probability of error becomes smaller. The broken lines show the threshold value A where r is larger than its minimum value by 20%, while the chain line shows the value of r in the case where the value of A deviates from the optimum value by ±20%. Therefore, with the variation of A, it is seen that the value of r does not change much near its optimum value. Furthermore, though r becomes larger with w_0, the increment of r is much less than that of w_0, where w_0 is proportional to the duration of the decision-applying period T_a. Accordingly, it will also be realized that the longer this duration becomes, the less the expectation of the total loss becomes. In fact, however, this duration will be limited to a certain magnitude in order to check the necessity of the renewal of the decision based upon the slow variation of the probability density function. When $u \neq 1$ in Eq.(29), B cannot explicitely be expressed by A. Then, the solution of A and B obtained by numerical calculation are shown in Figs.4b and 4c in the case of u=0.1 and 10. Now, as it can be seen from Eq.(28), the risk W_{10} in the case of u=0.1 is less than that in the case of u=1 (in Fig.4a) for the same value of w_0. It could generally be said that, the less the risk due to a decision becomes, the

larger error probability can be taken for the same expectation of the total loss. Hence the larger value of β in the former case can be taken than that in the latter case. Thus it may be understood from Eq.(20) that the value of B in Fig.4b is larger than that in Fig.4a and conversely the value of A in Fig.4b is smaller than that in Fig.4a. Similarly the characteristics of A in Fig.4c to that in Fig.4a can be explained. When the incorrect decision can be corrected as mentioned above, the graph $u=W_{10}/W_{01}>1$ will be useful.

6. Simulation

In order to simplify the simulation procedures, the following is assumed.

(1) The operating point is to be always at the optimum or suboptimum point by means of the local optimization.

(2) The difference between the PI of the optimum point and that of the suboptimum point is unity for each of n classes ($n=1,\cdots,10$).

(3) The distribution function of PI samples is a normal one having the same known dispersion σ^2 for each class.

(4) y' and d is given by

$$y'=y^*+d, \quad d=|\theta_1-\theta_0|/2 \tag{34}$$

(5) The risk due to incorrect decision is

$$W_{01}=W_{10}=2dm \tag{35}$$

The simulation procedures is,

(1) the class number j of the operating point is determined by generating the uniform random number, and thereby the classified value of PI y_j^* is determined. Next, the table of θ at the opposite optimum point is looked up, and then the PI sample y_j at the opposite optimum point is determined as follows;

$$v=F(y_j)=\int_{-\infty}^{y_j} f(y,\theta)dy \tag{36}$$

where v is a uniform random number $0\leq v\leq 1$ and $f(y,\theta)$ is given by Eq.(23). Thereafter go to (2).

(2) (a)If the decision for the jth class has already been made, go to (3).

(b)If the decision for the jth class has not yet been made, add the global exploration cost C_1 to SC_1 ($SC_1+C_1\rightarrow SC_1$), calculate Z_j by using the value of y_j, and

(i) if the value of Z_j belongs to the domain for which a decision has not yet been made, go to (1).

(ii) if the value of Z_j belongs to the domain for which a decision has already been made, and

(α) if the decision is incorrect, $SW+|\theta_1-\theta_0|\times m\rightarrow SW$ and go to (1).

(β) if the decision is correct, go immediately to (1).

(3)(a) In the application of the decision, if no loss yields, go to (1).

(b) In the application of the decision, if any loss yields, and

(i) if this loss is based on an incorrect decision, SWA+L→SWA and go to (1)

(ii) if this is due to the dispersion, though the decision itself is correct, SDA+L→SDA and go to (1).

Here SC_1 and SW are the cumulative loss due to the global exploration and the incorrect decision respectively. Similarly SWA and SDA are both the cumulative loss due to the incorrect decision being applied.

Figures 5a and 5b show, in relation to time, the average cumulative loss based on decisions made over 60 examples resulted from the simulation. The procedures of the simulation is almost the same as the above, and the number of decision-applying is taken to be m=100. It can be seen that the average number of inspection (global exploration) and the value of SC_1 both increase with the value of A, while the risk SW due to the incorrect decision is reduced.

The optimum value of A minimizing the total loss SC_1+SW and the corresponding minimum loss r are theoretically obtainable from Eq.(28) and Fig. 4a: A=19.2, r=3.96, (w_0=25) when σ^2 =1 and A=3.9, r=2.1, (w_0=6.25) when σ^2=4. These results coincide qualitatively with those of the above simulation, where A=20, $(SC_1+SW)/E_{\theta_1}/C_1$=4.7 when σ^2=1 and A=5.0, $(SC_1+SW)/E_{\theta_1}/C_1$ =2.5 when σ^2=4. Figure 6, in relation to time, shows a simulation result of the average cumulative loss over 60 examples due to the decision-applying. The risk SW due to the decision-making (for m=100) is expected to be equal to the loss SWA due to the 100 incorrect decisions being applied. Correspondingly, the qualitative coincidence of both the quantities is shown in Table 1.

It can also be seen from Table 1 that the theoretical values of average inspection number and error probability obtained from Eqs.(21) and (19) coincide qualitatively with the simulation results. Now, when the optimum point is always searched by the global exploration without using the above decision procedures, the expectation of the loss resulted in the case where the opposite PI value is less than that of the operating point will probably be $C_1 \times m/2$ (=2×100/2), since the number of the classes having larger mean value of PI than that of the opposite optimum point and the number of classes having the smaller value of PI would in many cases equal to each other. On the other hand, when the above decision procedures are applied the expected loss can be seen from Fig.4a rC_1/E_{θ_1}=15.8 provided that σ^2=1, A=19.2 and that the loss due to the dispersion is neglected. The difference between

their losses increases with C_1 and m_1, and the application of decision procedures becomes more advantageous. This will also be true in the case where the PI value at the opposite optimum point is larger than that at the operating point, because the expectation of the opposite location has been recorded and hence only the local exploration cost C_2 which is much less than C_1 is, in most cases, required.

7. The loss due to the dispersion

Though the decision is correct in the above procedures, the loss SDA due to the dispersion of PI samples yields in the decision-applying. Now, when the decision $d_e=d_0$ is correct, the expected loss due to the dispersion in one exploration is

$$L_0=(\sqrt{2\pi}\sigma)^{-1}\int_{\theta_1}^{\infty}(y-\theta_1)\varepsilon^{-(y-\theta_0)^2/(2\sigma^2)}dy$$
$$=(\sigma/\sqrt{2\pi})\varepsilon^{-(\theta_1-\theta_0)^2/(2\sigma^2)}-\{(\theta_1-\theta_0)/2\}\{1-(2/\sqrt{\pi})\int_0^{(\theta_1-\theta_0)/(\sqrt{2}\sigma)}\varepsilon^{-y^2}dy\} \quad (37)$$

Similarly, when the decision $d_e=d_1$ is correct, and if only the difference between their PI values is considered, the loss due to the dispersion in one exploration is

$$L_1=(\sqrt{2\pi}\sigma)^{-1}\int_{-\infty}^{\theta_0}(\theta_0-y)\varepsilon^{-(y-\theta_1)^2/(2\sigma^2)}dy=L_0 \quad (38)$$

Since $L_0=L_1$, the optimum values of A and B that minimize the maximum value of the sum of R_0 and L_0 or R_1 and L_1 remain the same as the above. Though this doesn't strictly hold in the general case where L_1 has to be evaluated in some more complicated forms from a practical viewpoint, this will approximately hold provided that the value of σ is small or the local optimization cost is much less than the value $|\theta_1-\theta_0|$.

Now, the value of L is given by Eq.(37): L=0.08 for σ=1 and L=0.39 for σ=2 respectively, provided that $\theta_1-\theta_0=1$. They correspond qualitatively to the simulation result in Fig.6 where the incremental rate of SDA is L≈0.08 for σ=1 and L≈0.31 for σ=2 respectively.

8. Selection of characteristic parameters

In the preceding sections, the parameters characterizing the PI value at the optimum opposite point was only one parameter such as the PI value at the operating point and its kind was predetermined beforhand. However, it would be necessary to adopt more parameters as conditional values of the conditional probability in accordance with the increasing number of independent disturbances, and it is very important to choose the appropriate characteristic parameters from the many available ones. The average mutual entropy, as is well known, is a useful measure for choosing the characteristic parameter, but it will rather be an indirect one with respect to the discrimination

rate and the loss. Here a method of choosing the characteristic parameters through the dispersion of the conditional probability is simply examined.

Let us choose x_i $(i=1,2,\cdots)$ the ith chatacteristic parameter from measurable quantities at the operating point. And x_i is assumed to be classified into N_i classes. Denoting by $f_{j,x1\cdots xi}$ the frequency in the class to which the opposite PI value y_j belongs for the conditional value $x_1,\cdots,x_i$ the average of y_j is

$$\bar{y}_{x1\cdots xi}=\frac{\sum_j f_{j,x1\cdots xi}y_j}{\sum_j f_{j,x1,\ldots,xi}} \tag{39}$$

The dispersion is

$$\sigma_{x1\cdots xi}^2=\sum_j f_{j,x1\cdots xi}(y_j-\bar{y}_{x1\cdots xi})/\sum_j f_{j,x1\cdots xi} \tag{40}$$

Putting

$$f_{x1\cdots xs}=\sum_j f_{j,x1\cdots xi} \tag{41}$$

the expectation of the dispersion is

$$\sigma_i^2=\sum_{x1\cdots xi}\sum_j f_{j,x1\cdots xi}(y_j-\bar{y}_{x1\cdots xi})^2/\sum_{x1\cdots xi} f_{x1\cdots xi} \tag{42}$$

Now, it will be seen from Eqs.(25), (28) and (37) that the expectation of the loss in decision becomes smaller as the value of σ_i^2 decreases. Therefore the characteristic parameter x_i should be sequentially chosen so that the value of σ_i^2 is as small as possible. Let us suppose that the characteristic parameters up to the ith are determined to be adopted as the conditional value. If the (i+1)th characteristic parameter is furthermore chosen properly and adopted as the conditional value, the dispersion will become smaller and generally $\sigma_i^2/\sigma_{i+1}^2\geq 1$. On the other hand, the inspection number necessary to decide all classes will become roughly N_{i+1} times larger. Then, as a simple criterion regarding the adoption of the (i+1)th characteristic parameter, the following will be considered:

If

$$\sigma_i^2/\sigma_{i+1}^2>N_{i+1} \tag{43}$$

the (i+1)th characteristic parameter is to be adopted. Otherwise, it is not to be adopted.

9. Conclusion

Since the optimization problem of the control system with multi-peak performance index contains the various complicated factors to be solved, it is very difficult at the present stage to solve the whole of the problem systematically and furthermore optimally. This paper discusses mainly the application of the minimax sequential probability ratio test to the estimation of the opposite optimum point from the data at the operating point. Though the theory of the sequential probability ratio test is comparatively

systematic and almost completed, the systematic optimization in the practical problems will be difficult because it includes various unknown factors such as the form of probability density function, the dispersion and the value of d in Eq.(11). However, the deviation from the optimum due to the incorrect estimation of the above factors will not be so serious a matter from a practical viewpoint since the increment of loss due to the deviation from the optimum value is generally small as is already described in section 5.

The other important aspects of the problem to be solved are the efficient global exploration method and the selection of the characteristic parameters. The considerations in section 3 and 8 are only heuristic and a kind of a trial. More efficient and systematic procedures are expected to be developed.

Reference

1. D.C. Karnopp; Random search techniques for optimization problems. Automatica, Vol.1, pp.111-121, August, 1963.
2. H.J. Kushner; A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. Journal of Basic Engineering, Trans.ASME, Vol.86, Series D, No.1, pp.97-106, March, 1964.
3. G.J. McMurtry and K.S. Fu; A variable structure automaton used as a multimodal searching technique. IEEE Trans. on AC-11, No.3, pp.379-389, July, 1966.

Table 1. Comparison of the theoretical values (in parenthesis) and the results of simulation.

A	in case of $\sigma^2=1$			in case of $\sigma^2=4$	
	5	20	50	5	40
the average inspection number	3.567 (2.146)	6.133 (5.421)	8.132 (7.500)	11.95 (8.58)	32.60 (28.07)
the error probability	0.13 (0.17)	0.066 (0.048)	0.033 (0.020)	0.18 (0.17)	0.033 (o.024)
the loss due to incorrect decision	14.0 13.3	7.5 6.7	3.3 3.3	22.3 18.0	3.50 3.33

In the column of the loss, SWA: the upper
SW : the lower

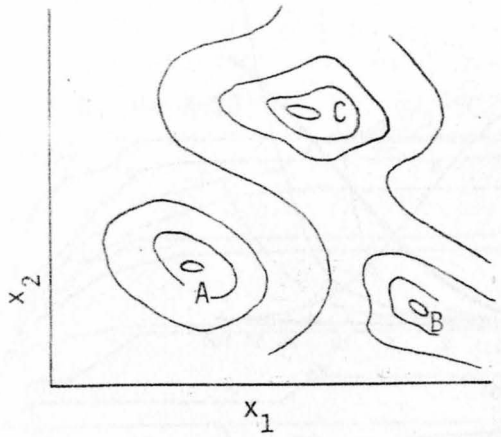


Fig.1. An example of multimodal response surface.

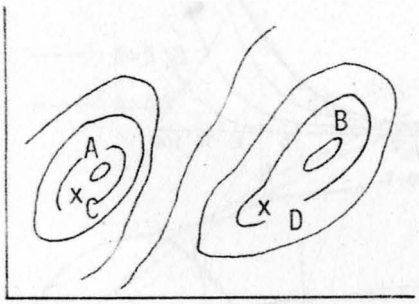


Fig.2. The peaks and their neighbourhood.

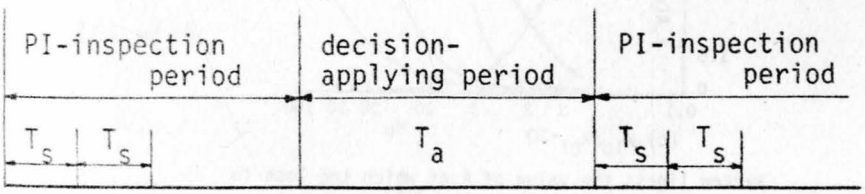
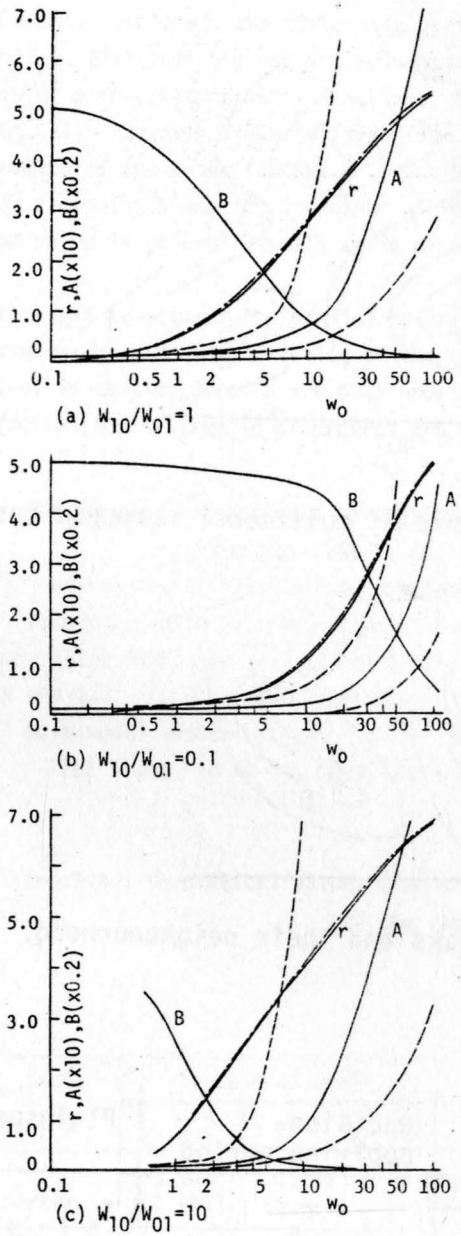


Fig.3. Alternation of PI-inspection period and the decision-applying period.



Broken lines: the value of A at which the loss in larger than r by 20%.

Chain lines; the value of r in the case where the value of A deviates from the optimum value by ± 20%.

Fig.4. The optimum values of A and B, and the loss r versus w_o that is proportional to the risk divided by the inspection cost.

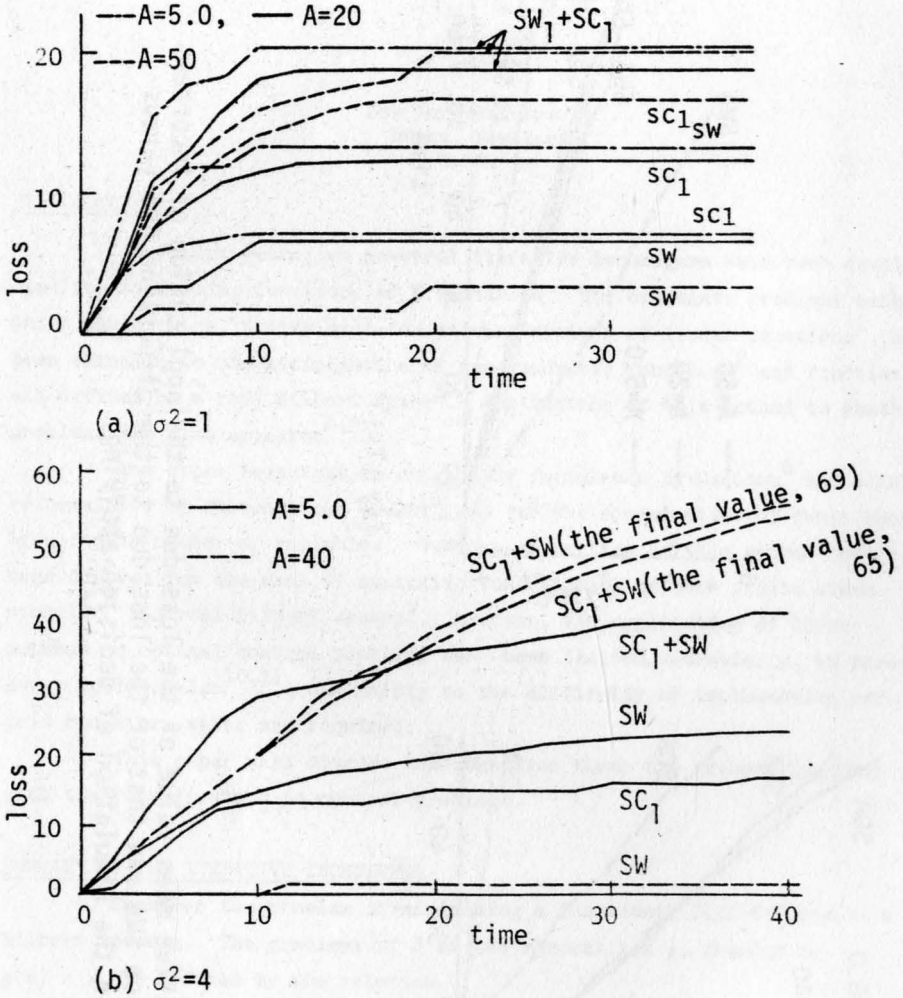
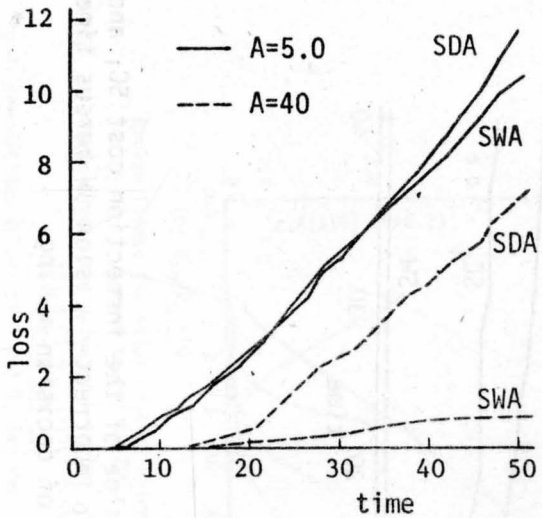
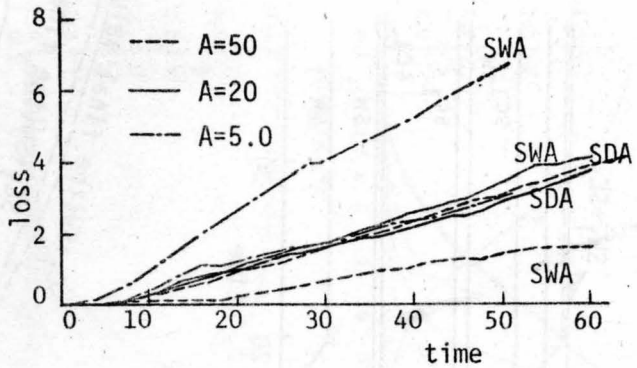


Fig.5. Cumulative mean value of the inspection cost SC_1 and that of risk due to incorrect decision SW versus time in the simulation of decision making.



(b) $\sigma^2=4$



(a) $\sigma^2=1$

Fig.6. Cumulative mean value of the loss due to the application of incorrect decision SWA and that of the loss due to the dispersion SDA versus time in the simulation of decision-applying.

A SURVEY OF SOME RECENT ITERATIVE TECHNIQUES FOR COMPUTING OPTIMAL CONTROL SIGNALS *

L. Horwitz †
P. Sarachik

New York University
Bronx, New York
U. S. A.

INTRODUCTION

In recent years, two powerful iterative techniques have been developed for minimizing functions of N variables. The conjugate gradient method, which was originally developed for solving systems of linear equations[1], has been extended to the minimization of non-quadratic functions[2] and functionals defined on a real Hilbert space[3]. Application of this method to control problems has also appeared[4,5].

The other technique, as originally formulated by Davidon[6] and later reformulated by Fletcher and Powell[7], was for the minimization of functions of a finite number of variables. Various properties of this method have been derived for the case of quadratic functionals, on both finite dimensional[7,8] and real Hilbert spaces[9]. However, the application of these methods to optimal control problems has been limited, essentially, to parameter optimization[10,11], due mainly to the difficulty of implementing certain operators which are required.

This paper will discuss and summarize these two methods together with their application to control problems.

SUMMARY OF THE ITERATIVE TECHNIQUES

Consider the problem of minimizing a functional $J[\underset{\sim}{x}]$ defined on a Hilbert space χ. The gradient of J at the element $\underset{\sim}{x} \in \chi$, denoted by $\underset{\sim}{g}(\underset{\sim}{x}) \in \chi$, is defined by the relation

$$\frac{\partial}{\partial \varepsilon} J[\underset{\sim}{x}+\varepsilon \underset{\sim}{h}]\Big|_{\varepsilon=0} = \langle \underset{\sim}{g}(\underset{\sim}{x}), \underset{\sim}{h} \rangle \qquad (1)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. It is assumed that the gradient element $\underset{\sim}{g}(\underset{\sim}{x})$ as defined above exists for all elements $\underset{\sim}{x} \in \chi$.

*This research was sponsored in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant AFOSR-747-66.

†Presently at Bell Telephone Laboratories, Whippany, New Jersey, U.S.A.

Denote by $\underset{\sim}{x}^*$ the element that minimizes J. Then in general the iteration schemes attempt to produce a sequence $\{\underset{\sim}{x}^i\}$ which converges to $\underset{\sim}{x}^*$ by using the iteration rule

$$\underset{\sim}{x}^{n+1} = \underset{\sim}{x}^n + \alpha_n \underset{\sim}{s}^n .$$

An initial element $\underset{\sim}{x}^o \in \chi$ is chosen arbitrarily. At each subsequent step a descent direction $\underset{\sim}{s}^n$ is chosen (the way this is done defines the method used) and a step size α_n is determined such that

$$J[\underset{\sim}{x}^n + \alpha_n \underset{\sim}{s}^n] \leq J[\underset{\sim}{x}^n + \lambda \underset{\sim}{s}^n]$$

for all λ. This leads to the condition

$$J[\underset{\sim}{x}^{n+1}] \leq J[\underset{\sim}{x}^n] \leq \ldots \leq J[\underset{\sim}{x}^o].$$

In Davidon's method the descent direction is chosen as

$$\underset{\sim}{s}^n = -H^n \underset{\sim}{g}^n$$

where H^n is an operator defined by

a. $$H^n = H^{n-1} + C^{n-1} \urcorner B^{n-1}$$

b. $$B^{n-1} = \frac{H^{n-1}\underset{\sim}{y}^{n-1} \langle H^{n-1}\underset{\sim}{y}^{n-1}|}{\langle \underset{\sim}{y}^{n-1}, H^{n-1}\underset{\sim}{y}^{n-1}\rangle}$$

c. $$C^{n-1} = \frac{\underset{\sim}{\sigma}^{n-1}\langle \underset{\sim}{\sigma}^{n-1}|}{\langle \underset{\sim}{\sigma}^{n-1}, \underset{\sim}{y}^{n-1}\rangle}$$

d. $$\underset{\sim}{y}^{n-1} = \underset{\sim}{g}^n - \underset{\sim}{g}^{n-1}$$

e. $$\underset{\sim}{\sigma}^{n-1} = \alpha_{n-1} \underset{\sim}{s}^{n-1}$$

The dyadic notation for operators, as defined in references [12 and 13] has been used, and the initial value of the H operator is chosen to be any strongly positive, bounded, linear, self adjoint operator in χ.

In the conjugate gradient method, $\underset{\sim}{s}^n$ is generated using the iteration

$$\underset{\sim}{s}^n = -\underset{\sim}{g}^n + \beta_{n-1}\underset{\sim}{s}^{n-1} \text{ with } \underset{\sim}{s}^o = -\underset{\sim}{g}^o,$$

where

$$\beta_{n-1} \triangleq \frac{||\underset{\sim}{g}^n||^2}{||\underset{\sim}{g}^{n-1}||^2}. \qquad (2)$$

Various properties of these schemes, along with convergence proofs, when J is a quadratic and strongly positive functional can be found in references [3 and 9]. In particular, for quadratic functionals the Davidon method gives [9]

$$\underset{\sim}{s}^n = -||\underset{\sim}{s}^n||^2 \sum_{i=o}^{n} \frac{H^o \underset{\sim}{g}^i}{\langle \underset{\sim}{g}^i, H^o \underset{\sim}{g}^i\rangle} \qquad (3)$$

$$\langle \underset{\sim}{s}^i, A\,\underset{\sim}{s}^j \rangle = 0 \text{ and } \langle \underset{\sim}{g}^i, H^o \underset{\sim}{g}^j \rangle = 0 \quad \text{for } i \neq j,$$

and the conjugate gradient method gives[3]

$$\beta_{n-1} = \frac{\langle \underset{\sim}{g}^n, A\underset{\sim}{s}^{n-1} \rangle}{\langle \underset{\sim}{s}^{n-1}, A\underset{\sim}{s}^{n-1} \rangle} \tag{4}$$

$$\underset{\sim}{s}^n = -||\underset{\sim}{g}^n||^2 \sum_{i=0}^{n} \frac{\underset{\sim}{g}^i}{||\underset{\sim}{g}^i||^2}$$

$$\langle \underset{\sim}{s}^i, A\,\underset{\sim}{s}^j \rangle = 0 \text{ and } \langle \underset{\sim}{g}^i, \underset{\sim}{g}^j \rangle = 0 \text{ for } i \neq j,$$

where A is defined by the relation

$$\frac{\partial}{\partial \varepsilon} \underset{\sim}{g}(\underset{\sim}{x}+\varepsilon \underset{\sim}{h})\Big|_{\varepsilon=0} = A\underset{\sim}{h}$$

i.e. A is the second derivative operator of J at $\underset{\sim}{x}$ and is independent of $\underset{\sim}{x}$ for quadratic functionals. Note that for quadratic J

$$A\underset{\sim}{s}^{n-1} = \frac{1}{\varepsilon}\{ \underset{\sim}{g}(\underset{\sim}{x}^n+\varepsilon \underset{\sim}{s}^{n-1}) - \underset{\sim}{g}(\underset{\sim}{x}^n) \}$$

whereas the relation is approximately true for small ε when J is nonquadratic. Thus if H^o is chosen as the identity operator, then for quadratic functionals the Davidon and conjugate gradient methods are identical.

Note that (4) suggests a generalization of the conjugate gradient method for non-quadratic functionals which reduces to (2) in the quadratic case. This generalization is discussed in reference[5] while the form given in (2) is discussed in reference.[2,4] Similarly (3) suggests an alternate iteration method which reduces to the Davidon Method for quadratic functions. That is. the descent directions could be generated via

$$\underset{\sim}{s}^n = -H^o \underset{\sim}{g}^n + \gamma_{n-1}\, \underset{\sim}{s}^{n-1} \tag{5}$$

where

$$\gamma_{n-1} = \frac{\langle \underset{\sim}{g}^n, H^o \underset{\sim}{g}^n \rangle}{\langle \underset{\sim}{g}^{n-1}, H^o \underset{\sim}{g}^{n-1} \rangle}.$$

Then it turns out that

$$\underset{\sim}{s}^n = -\langle \underset{\sim}{g}^n, H^o \underset{\sim}{g}^n \rangle \sum_{i=0}^{n} \frac{H^o \underset{\sim}{g}^i}{\langle \underset{\sim}{g}^i, H^o \underset{\sim}{g}^i \rangle}$$

and it follows that for quadratic functionals the descent directions $\underset{\sim}{s}^n$ are proportional to those given by the Davidon method. This iteration method has in fact been presented as a generalization of the original conjugate gradient method.[1]

APPLICATIONS TO PROBLEMS OF OPTIMAL CONTROL

Consider the following problem of finding the u^* which minimizes

the cost (which is assumed to have a minimum)

$$J[\underset{\sim}{u},t_m] = \int_{t_o}^{tm} f_o(\underset{\sim}{x}(t), \underset{\sim}{u}(t),t)\, dt \tag{6}$$

subject to the side constraint

$$\dot{\underset{\sim}{x}} = \underset{\sim}{f}(\underset{\sim}{x}(t), \underset{\sim}{u}(t), t) \tag{7}$$

with $\underset{\sim}{x}(t_o)=\underset{\sim}{x}_o$, $\underset{\sim}{x}_o$ and t_o given and t_m either fixed or free. Here $\underset{\sim}{x}(t)$ is an n-vector and $\underset{\sim}{u}(t)$ is an r-vector. By introducing an n-vector of Lagrange multipliers (or adjoint variables) $\underset{\sim}{p}(t)$, one may show that the gradient of J with respect to $\underset{\sim}{u}$, denoted by $\nabla_{\underset{\sim}{u}} J$, is given by[14]

$$\nabla_{\underset{\sim}{u}} J = \underset{\sim}{p}^T \frac{\partial \underset{\sim}{f}}{\partial \underset{\sim}{u}} + \frac{\partial f_o}{\partial \underset{\sim}{u}} \tag{8}$$

where $\underset{\sim}{p}(t)$ satisfies the differential equation

$$\dot{\underset{\sim}{p}}(t) = -\left(\frac{\partial f_o}{\partial \underset{\sim}{x}}\right)^T - \left(\frac{\partial \underset{\sim}{f}}{\partial \underset{\sim}{x}}\right)^T \underset{\sim}{p}(t) \tag{9}$$

with $\underset{\sim}{p}(t_m)=0$, and $\underset{\sim}{x}(t)$ satisfies equation (7). Thus, for the case when the terminal time t_m is fixed, an iterative solution might proceed as follows:

1. For n = 0 guess an initial control $\underset{\sim}{u}^o$.
2. Integrate (7) forward to obtain $\underset{\sim}{x}^n$.
3. Integrate (9) backward to obtain $\underset{\sim}{p}^n$ and at the same time calculate $(\nabla_{\underset{\sim}{u}} J)^n$ as given in (8).
4. Operate on $(\nabla_{\underset{\sim}{u}} J)^n$ to obtain a descent direction $\underset{\sim}{s}^n$ as discussed earlier.
5. Form
$$\underset{\sim}{u} = \underset{\sim}{u}^n + \alpha\, \underset{\sim}{s}^n$$
and search for α_n such that $J[\underset{\sim}{u}^n+\alpha_n \underset{\sim}{s}^n] \leq J[\underset{\sim}{u}^n+\alpha \underset{\sim}{s}^n]$.
6. Set $\underset{\sim}{u}^{n+1} = \underset{\sim}{u}^n + \alpha_n\, \underset{\sim}{s}^n$.
7. Go to step 2 and continue with n=n+1.

The case of a variable terminal time is handled by using the gradient of J with respect to t_m. This is given by

$$\nabla_{t_m} J = f_o(\underset{\sim}{x}(t_m), \underset{\sim}{u}(t_m), t_m).$$

Thus, by considering $\underset{\sim}{z} = [\underset{\sim}{u},t_m]^T$ one may iterate on $\underset{\sim}{z}$ to obtain the minimizing element $\underset{\sim}{z}^*$, i.e.,

$$\underset{\sim}{z}^{n+1} = \underset{\sim}{z}^n + \alpha_n \underset{\sim}{s}^n \tag{10}$$

where the descent direction is

$$\underset{\sim}{s}^n = [s_1^n, s_2^n]^T$$

Since u^n and s_1^n are both defined over the time interval $[t_o, t_m^n]$ then so is u^{n+1} by equation (10). But during the same iteration t_m^n is updated to t_m^{n+1}, and if $t_m^{n+1} > t_m^n$ then u^{n+1} is not defined over the time interval $[t_m^n, t_m^{n+1}]$. A procedure usually adopted to handle this difficulty is to arbitrarily set $\underset{\sim}{u}^{n+1}(t) = \underset{\sim}{u}^{n+1}(t_m^n)$, $t_m^n \leq t \leq t_m^{n+1}$. The iterative solution would then proceed as follows

1. For n = 0 guess an initial $\underset{\sim}{u}^o$ and t_m^o.
2. Integrate (7) forward to obtain $\underset{\sim}{x}^n$ and $(\nabla_{t_m} J)^n$.
3. Integrate (9) backwards to obtain $\underset{\sim}{p}^n$ and $(\nabla_u J)^n$ as given in (8).
4. Operate on $\underset{\sim}{z}^n = [\underset{\sim}{u}^n, t_m^n]$ to obtain a descent direction $\underset{\sim}{s}^n$ as discussed previously.
5. Form

   $$\underset{\sim}{z} = [\underset{\sim}{u}, t_m]^T = \underset{\sim}{z}^n + \alpha \underset{\sim}{s}^n$$

   and search for α_n such that $J[\underset{\sim}{z}^n + \alpha_n \underset{\sim}{s}^n] \leq J[\underset{\sim}{z}^n + \alpha \underset{\sim}{s}^n]$, arbitrarily setting $\underset{\sim}{u}(t) = \underset{\sim}{u}(t_m^n)$ for $t_m^n \leq t \leq t_m$ when $t_m > t_m^n$.
6. Set $\underset{\sim}{z}^{n+1} = \underset{\sim}{z}^n + \alpha_n \underset{\sim}{s}^n$.
7. Go to step 2 and continue with n = n+1.

Another procedure for the variable terminal time problem can be developed as follows: Instead of guessing an initial t_m^o and iterating on t_m and $\underset{\sim}{u}$ simultaneously until J is minimized with respect to $\underset{\sim}{u}$ and t_m, one may attempt to choose t_m at every iteration to satisfy $J[\underset{\sim}{u}^n, t_m^n] \leq J[\underset{\sim}{u}^n, t_m]$ for all t_m. This is done in step 2 where the forward integration is now terminated at a time t_m^n such that $\nabla_{t_m} J = 0$, with a check to see that J is minimized with respect to t_m for $\underset{\sim}{u} = \underset{\sim}{u}^n$. Note that using the condition $\nabla_{t_m} J = 0$ to define t_m^n is not sufficient since this may result in choosing t_m^n such that $J[\underset{\sim}{u}^n, t_m^n]$ is a maximum or saddle value. In steps 4-6, $\underset{\sim}{u}$ is then used instead of $\underset{\sim}{z}$.

A third procedure for the free terminal time case is to solve a sequence of fixed terminal time problems. This is done by setting t_m to t_m^n, solving the fixed terminal time problem for $\underset{\sim}{u}_n^*$ and $\underset{\sim}{x}_n^*$, and then updating the terminal time using a descent method and $\nabla_{t_m} J[\underset{\sim}{u}_n^*, t_m^n]$.

A different approach can also be utilized to solve the control problem stated. Suppose that, in addition to the differential side constraints given in (7), $\underset{\sim}{u}(t)$ is required to be a piecewise constant time

function, i.e.,

$$\underline{u}(t) = \underline{h}_j \quad \text{for} \quad t_{j-1} \leq t < t_j, \quad j=1,\ldots,m \tag{11}$$

where the amplitudes $\underline{h}_j$ and switching times t_j are to be chosen to minimize J subject to the condition

$$t_{j-1} \leq t_j, \qquad j=1,\ldots,m \tag{12}$$

Thus, when m is a fixed finite number, the original problem is reduced to one of minimizing a function of a finite number of variables. In other words,

$$\begin{aligned} J[\underline{u},t_m] &= J(\underline{h}_1, \ldots, \underline{h}_m; t_1, \ldots, t_m) \\ &= \sum_{j=1}^{m} \int_{t_{j-1}}^{t_j} f_o(\underline{x}(t),\underline{h}_j,t)dt \end{aligned}$$

To solve this problem iteratively, a penalty function term is introduced into J to eliminate the inequality constraints (12) on the t_j's. Let

$$\begin{aligned} f_{oj} &= f_o(\underline{x}(t),\underline{h}_j,t) \\ \underline{f}_j &= \underline{f}(\underline{x}(t),\underline{h}_j,t) \\ q_j &\triangleq 1(t_j-t_{j-1}) - k_j \operatorname{sgn}(f_{oj})\, 1(t_{j-1}-t_j) \end{aligned}$$

where

$$1(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } t > 0 \end{cases},$$

the $k_j > 1$ are predetermined penalty constants and

$$\operatorname{sgn}(f) = \begin{cases} -1 & \text{for } f < 0 \\ 1 & \text{for } f > 0 \end{cases}.$$

Now consider $\hat{J}(\underline{h}_1, \ldots, \underline{h}_m; t_1, \ldots, t_m) = \sum_{j=1}^{m} \int_{t_{j-1}}^{t_j} f_{oj} q_j dt.$

It is seen that when $t_j > t_{j-1}$ for all j, we have $\hat{J} = J$. Conversely, if $t_j < t_{j-1}$ for any J, then $\hat{J} \geq J$ for a suitable choice of the penalty constants. Thus the problem to be considered is that of minimizing $\hat{J}$, with an appropriate choice of penalty constants, subject to the given differential side constraints and initial conditions.

It can be shown that the gradient of J with respect to $\underline{h}_j$ and t_j is given by

$$\nabla_{\underline{h}_j} \hat{J} = \int_{t_{j-1}}^{t_j} \left\{ q_j \frac{\partial f_{oj}}{\partial \underline{h}_j} + \underline{p}^T \frac{\partial \underline{f}_j}{\partial \underline{h}_j} \right\} dt \tag{13}$$

$$j = 1, \ldots, m$$

$$\nabla_{t_j} \hat{J} = f_{oj} q_j - f_{oj+1} q_{j+1} + \underline{p}^T [\underline{f}_j - \underline{f}_{j+1}] \tag{14}$$

$$t=t_j, \; j=1,\ldots,m-1$$

$$\nabla_{t_m} \hat{J} = f_{om} q_m \quad \text{for } t=t_m \tag{15}$$

where

$$\dot{\underline{x}}(t) = \underline{f}_j \qquad t_{j-1} \leq t < t_j,\ j=1,\ \ldots,\ m \tag{16}$$

and $\underline{x}(t_o) = \underline{x}_o$

$$\dot{\underline{p}}^T(t) = -q_j \frac{\partial f_{oj}}{\partial \underline{x}} - \underline{p}^T \frac{\partial \underline{f}_j}{\partial \underline{x}} \qquad t_{j-1} \leq t < t_j,\ j=1,\ldots,m \tag{17}$$

and $\underline{p}(t_m) = \underline{0}$.

Thus a computational solution might proceed as follows:

1. Guess an initial set of amplitudes and switching times $\underline{h}_1^o,\ \ldots,\ \underline{h}_m^o, t_1^o,\ \ldots,\ t_{m-1}^o$. If t_m is specified, set t_m^o to the specified value.
2. Integrate equation (16) forward to obtain $\underline{x}^n$.
3. Integrate equation (17) backward to obtain $\underline{p}^n$.
4. Calculate $(\nabla_{\underline{h}_j} \hat{J})^n$, $(\nabla_{t_j} \hat{J})^n$ for j=1, ..., m, from equations (13) - (15). Set $(\nabla_{t_m} J)^n = 0$ if the terminal time is specified so that t_m^n is held fixed at t_m^o.
5. Operate on $\underline{g} = [\nabla_{\underline{h}_1} \hat{J},\ldots,\nabla_{\underline{h}_m} \hat{J},\ \nabla_{t_1} \hat{J},\ldots,\nabla_{t_m} \hat{J}]$ to obtain a new descent direction $\underline{s}^n$ and step size α_n as discussed earlier.
6. Update

$$\underline{y} \triangleq [\underline{h}_1,\ldots,\underline{h}_m,t_1,\ldots t_m]$$

by

$$\underline{y}^{n+1} = \underline{y}^n + \alpha_n \underline{s}^n$$

7. Go to step 2 and continue with n = n+1.

Note that there is no theoretical difficulty encountered for the free terminal time case in this procedure.

A computational technique, based on this procedure, for generating optimal piecewise constant control signals may thus be formulated for the control problem stated in equations (6) and (7). Here we may seek out an optimal piecewise constant control signal with only a few switching times and use this as an initial guess for a piecewise constant control signal with more switching times, thereafter repeating the process until little improvement is obtained. This technique has been applied to several examples[11].

EXAMPLES

Two examples illustrating the previously discussed techniques are next presented.

Example 1

Here a control signal that minimizes a quadratic cost over a fixed time interval is to be found, where the system equations are nonlinear. Specifically

$$\dot{x}_1 = (1-x_2^2)x_1 - x_2 + u$$

$$\dot{x}_2 = x_1$$

$$x_1(0) = 0\,,\ x_2(0) = 1$$

and

$$J = \int_0^5 [x_1^2 + x_2^2 + u^2]dt$$

The solution to this example is given by Merriam[15]. The 5, 10 and 20 step optimal piecewise constant control signals have been computed using the Davidon method as defined in equations (a)-(e) with H^0 taken as the identity operator, and are plotted along with Merriam's solution in Figure 1.

The cost obtained by Merriam (here assumed to be the optimal cost) is 2.8670 while the costs obtained using the 5, 10 and 20 step solutions are 2.9848, 2.8893 and 2.8732. The percent deviations between these and the optimal cost, defined as

$$\%\ \text{deviation} = \frac{\text{cost - optimal cost}}{\text{optimal cost}} \times 100$$

are 4.1088%, .7782% and .2163% respectively, while the number of iterations required are 26, 22 and 17. Since the percent deviation for the 20 step case is quite small it did not seem necessary to continue the iterations using a piecewise constant control signal with more segments.

The initial guess used in obtaining the 5 step solution was an identically zero control signal with switching times at 1, 2, 3 and 4 seconds. The 5 step solution was then used as the initial guess for the 10 step solution and the 10 step solution was then used as the initial guess for the 20 step solution.

For a comparison of convergence rates, the example was also run using the conjugate gradient method with β_i chosen using (2). Figure 2 summarizes the convergence rates for the 20 step iterations.

Example 2

In this example, a minimum time problem is considered. Here we have

$$\dot{x}_1 = x_3\sin(u)$$
$$\dot{x}_2 = x_3\cos(u)$$
$$\dot{x}_3 = 32\sin(u)$$
$$x_i(0) = 0 \qquad i = 1,2,3$$
$$x_1(t_m) = 32$$

and

$$J = t_m$$

The solution to this problem is given in reference[16] and is

$$u(t) = 1.5708 - .88622t \qquad 0 \leq t \leq t_m$$
$$t_m = 1.7725$$

In order to put this problem into the form of equations (6) and (7), the terminal constraint on x_1 is eliminated by introducing a penalty function term in J. The cost functional now considered is

$$J = t_m + \tfrac{1}{2}K[x_1(t_m)-32]^2 = \int_0^{t_m}[1+\tfrac{1}{2}K\frac{d}{dt}(x_1-32)2]dt + \tfrac{1}{2}K(32)^2$$
$$= \int_0^{t_m}[1+K(x_1-32)x_3\sin(u)]dt + 512K$$

The modified problem was solved using Davidon's method as defined in equations (a) - (e) with H^o taken as the identity operator, and a 5 step piecewise constant control signal. The initial control signal was assumed identically zero with switching times at .4, .8, 1.2, and 1.6 seconds. A terminal time of 2 seconds was also assumed initially. The solution was found by iterating on the penalty constant for K = 10 and 20. For K = 20 the terminal value of x_1 was 32.000, the terminal time was 1.8590 and the total number of iterations required was 12. The solution is shown in Figure 3 along with the optimal control.

CONCLUSION

This paper has presented two powerful iterative techniques for minimization. Various properties which have been derived elsewhere were discussed and the application of these methods to problems of optimal control has been indicated. Two examples were included to illustrate these techniques.

Further discussion of these methods, as to computer storage, relative computation time and numerical roundoff, may be found in References 5 and 10. It should be noted that one may use approximating curves for $\underline{u}(t)$ other than the piecewise constant time functions considered in equation (11). For example, in problems whose control is continuous, it would be reaonsable to approximate using piecewise linear segements with continuity imposed. Such a procedure would result in only one more parameter than the one using piecewise constant segments.

REFERENCES

1. Hestenes, M.R., "The Conjugate Gradient Method for Solving Linear Systems," Proc. of Symposium in Applied Mathematics, Vol. - VI, Num. Anal., p. 83, 1956.
2. Fletcher, R. and Reeves, C.M., "Function Minimization by Conjugate Gradients," The Computer Journal, vol. 7, p. 149, 1964.
3. Antosiewicz, H.A. and Rheinboldt, W.F., "Numerical Analysis and Functional Analysis," Chapter 14 in the book "Survey of Numerical Analysis," J. Todd, editor, published by McGraw-Hill Company, 1962.
4. Lasdon, L.S., Mitter, S.K. and Waren, A.D., "The Conjugate Gradient Method for Optimal Control Problems," IEEE Transactions on Automatic Control, vol. AC-12, no. 2, p. 132, 1967.
5. Sinnott, J.F. and Luenberger, D.G., "Solution of Optimal Control Problems by the Method of Conjugate Gradients," JACC preprint, 1967.
6. Davidon, W.C., "Variable Metric Method for Minimization," A.E.C. Research and Development Report, ANL-5990, 1959.
7. Fletcher, R. and Powell, M.J.D., "A Rapidly Convergent Descent Method for Minimization," The Computer Journal, vol. 6, p. 163, 1964.
8. Myers, G.E., "Properties of the Conjugate Gradient and Davidon Methods," Journal of Optimization Theory and Applications, to appear.
9. Horwitz, L.B. and Sarachik, P.E., "Davidon's Method in Hilbert Space," S.I.A.M. Journal on Applied Math. vol. 16, p. 676, 1968.
10. Kelley, H.J. and Myers, G.E., "Conjugate Direction Methods for Parameter Optimization," Presented at the 18th Congress of the International Astronautical Federation, Belgrade, Yugoslavia, Sept. 24-30, 1967.
11. Horwitz, L.B. and Sarachik, P.E., "A Computational Technique for Calculating the Optimal Control Signal for a Specific Class of Problems," submitted for presentation at the Sixth Annual Allerton Conference, October 1968.

12. Porter, W.A., "Modern Foundations of Systems Engineering," published by The Macmillan Company.

13. Friedman, B., "Principles and Techniques of Applied Mathematics," published by Wiley Company.

14. Horwitz, L.B., "An Investigation of Iteration Techniques in Function Spaces and Applications to Optimal Control Problems," Ph.D. THESIS, New York University, Dept. of E.E., June 1968.

15. Merriam, C.W., "Optimization Theory and The Design of Feedback Control Systems," McGraw-Hill, Inc., 1964.

16. Bryson, A., Denham, W., Dreyfus, S., "Optimal Programming Problems with Inequality Constraints I: Necessary Conditions for Extremal Solutions," AIAAJ, vol. 1, pp. 2544-2550, 1963.

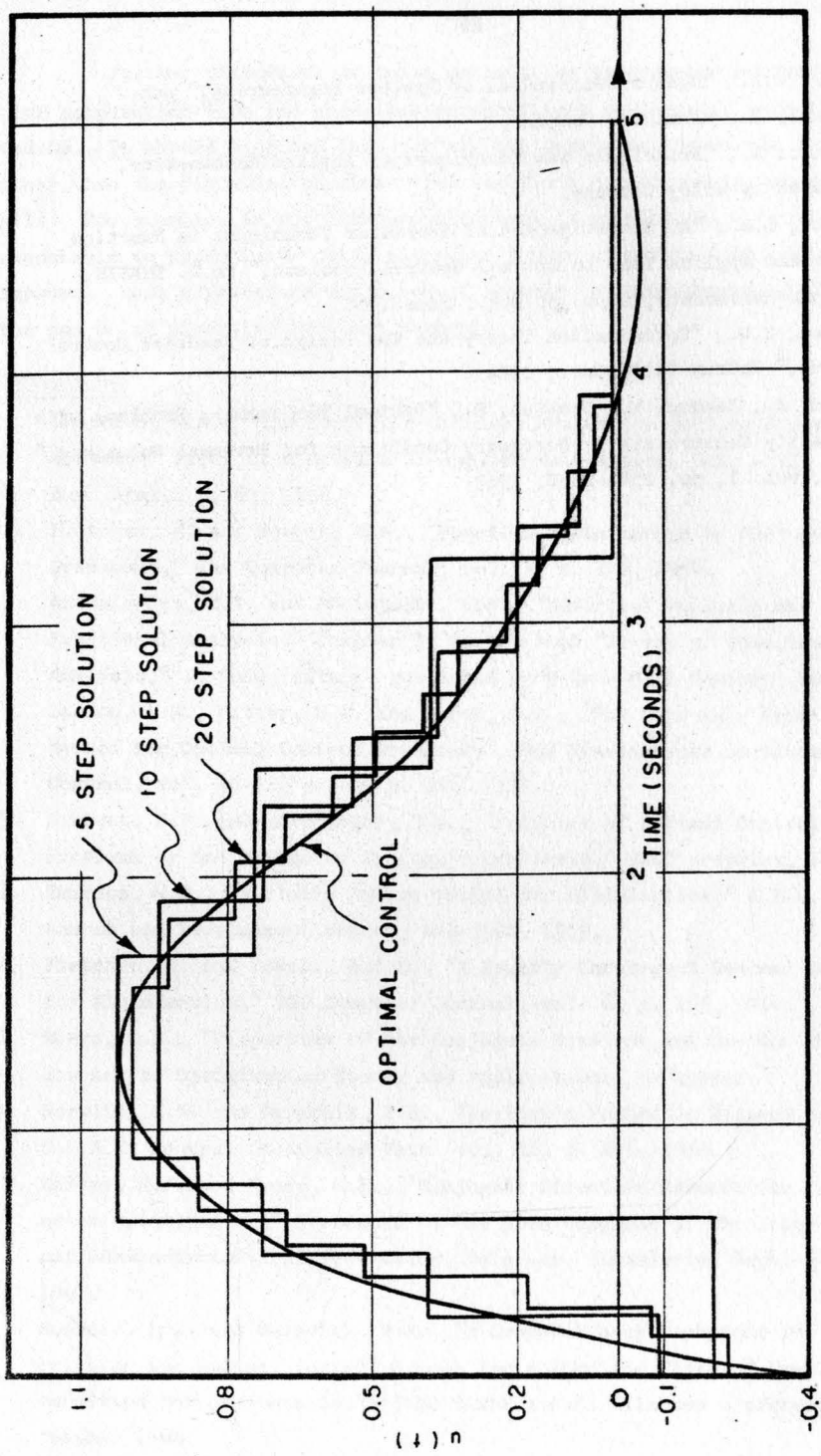

5 STEP SOLUTION
10 STEP SOLUTION
20 STEP SOLUTION
OPTIMAL CONTROL
TIME (SECONDS)
1.1
0.8
0.5
0.2
0
-0.1
-0.4
1
2
3
4
5

FIG. 1

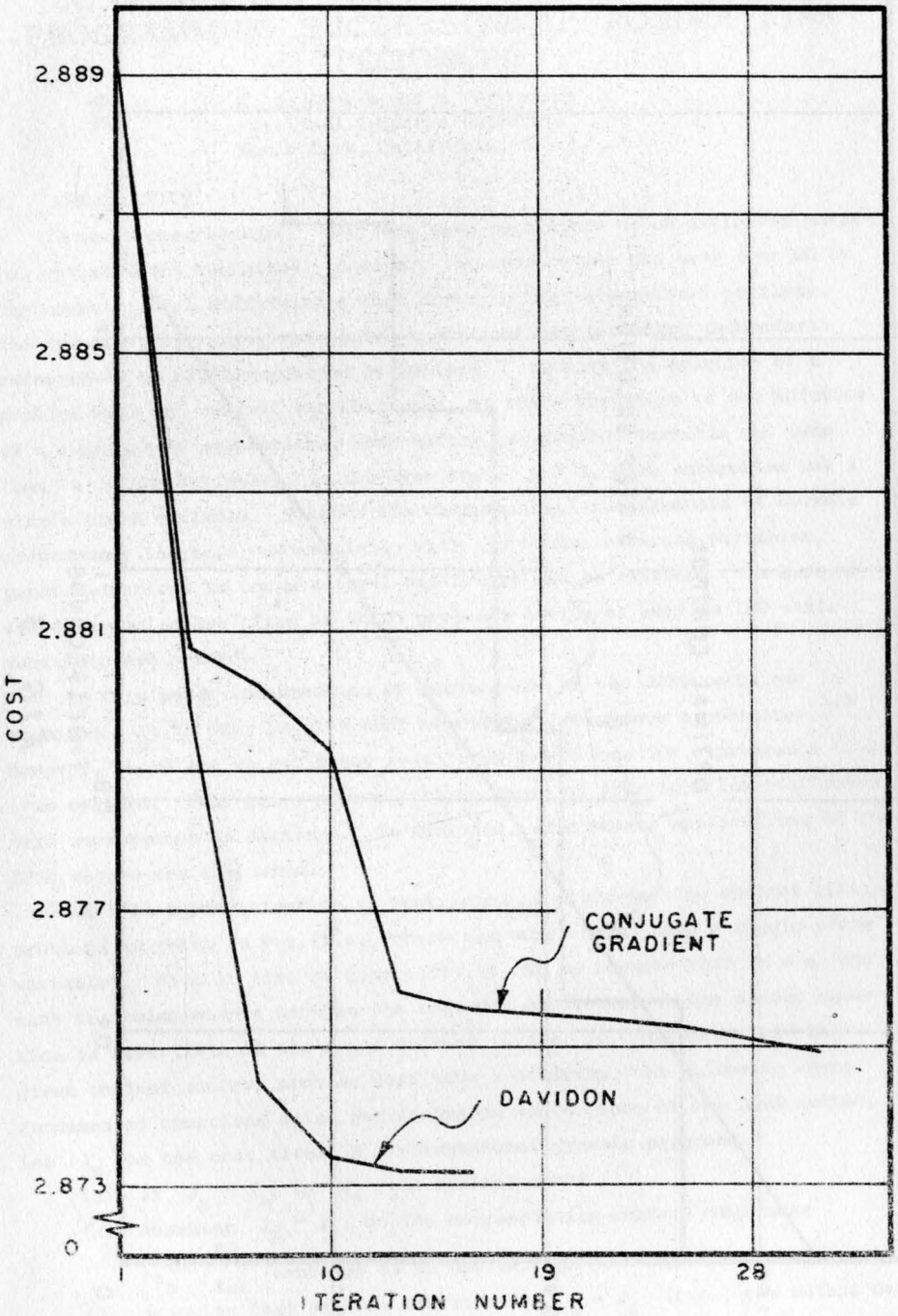

2.889
2.885
2.881
2.877
2.873
0
COST
1
10
19
28
ITERATION NUMBER
CONJUGATE GRADIENT
DAVIDON

FIG. 2

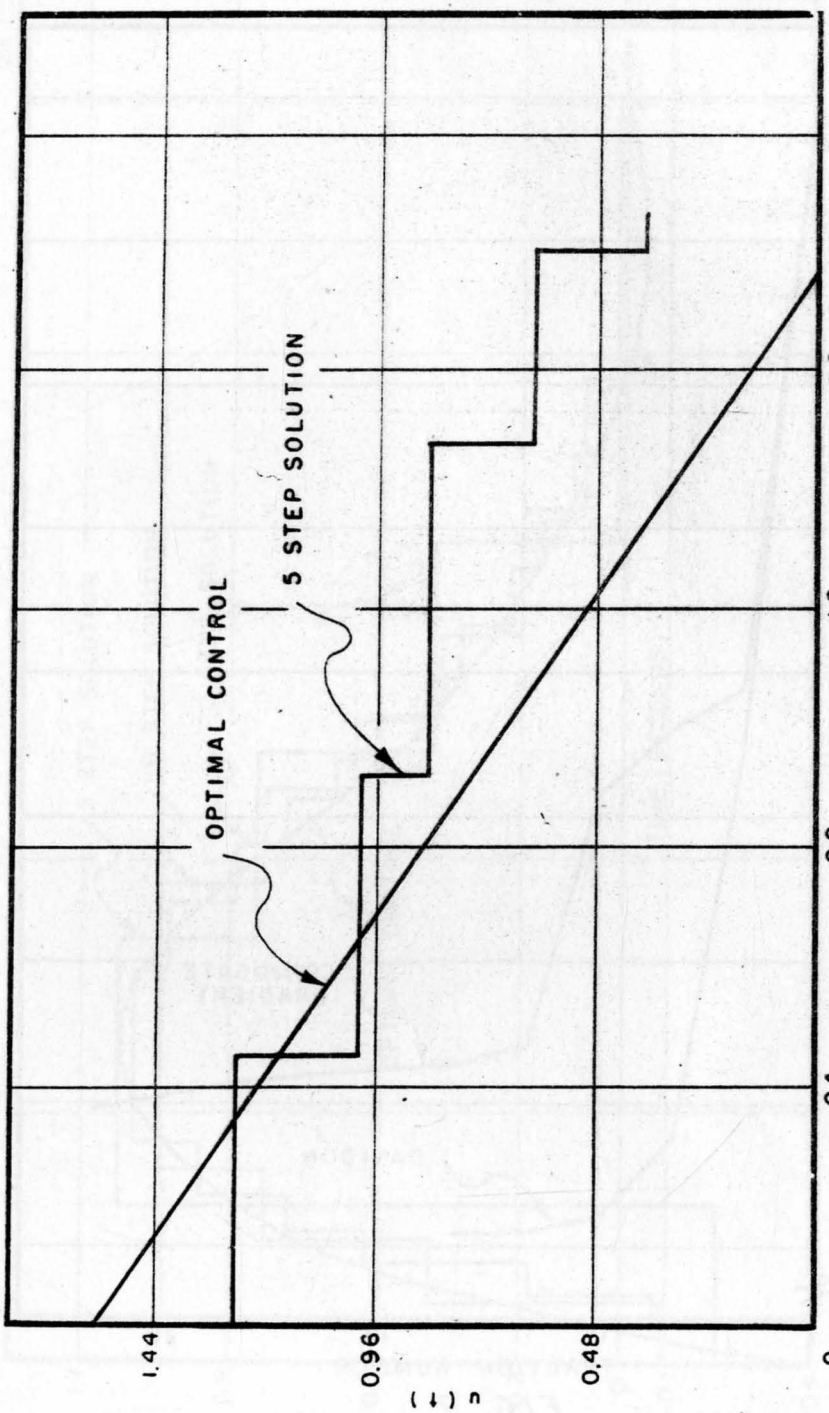

OPTIMAL CONTROL
5 STEP SOLUTION
1.44
0.96
0.48
0
u (t)
0.4
0.8
1.2
1.6
2.0
TIME (SECONDS)

FIG. 3

CONVERGENCE PROOFS FOR A DYNAMIC - -PROGRAMMING SUCCESSIVE-APPROXIMATIONS TECHNIQUE

A. J. Korsak and R. E. Larson
Stanford Research Institute
Menlo Park, California, U.S.A.

I INTRODUCTION

Dynamic programming[1,2] has long been recognized as an extremely powerful optimization technique. However, its application has been limited by the computational difficulties that occur in high-dimensional problems. The dynamic-programming successive-approximations technique (henceforth referred to as DPSA) suggested by Bellman[1,2] reduces the solution of a problem with m control variables and n state variables to the solution of a sequence of subproblems, each having one control variable and less than n state variables; in the case where $m \geq n$, each subproblem has a single state variable. Because the computational requirements of dynamic programming increase exponentially with the number of state variables, great reductions in computational difficulty can be obtained. In Refs. 3 and 4, examples are cited in which problems having as many as 100 state variables are solved.

In this paper the question of convergence of the DPSA technique is examined. It is easy to show that monotonic convergence is obtained.[1,2] However, there are no published results on conditions for convergence to a true optimum. This paper presents three cases of practical interest where this convergence is obtained. In addition, some recent applications of the DPSA method are discussed.

In this paper attention is restricted to discrete-time optimal control problems in which $m \geq n$ (i.e., where the subproblems have a single state variable). Without loss of generality it can be assumed that $m = n$ and that the relationship between the controls and states in the system equation is invertible in the sense defined in Sec. IV. Assuming that the given optimal control problem does have a solution, the following three fundamental questions arise pertaining to convergence of the DPSA method. Let J_N be the cost after N 1-dimensional dynamic programs.

(1) If $J_N \rightarrow J_\infty$, is J_∞ the optimal cost?

(2) Assuming $J_N \rightarrow J_\infty$, do the corresponding control sequences $\{u_k^{(N)}\}_{k=1}^{K}$ converge to an optimal control sequence?

(3) Assuming that at some iteration $J_{N+1} = J_N$ (i.e., the method can no longer reduce the cost at all), is J_N optimal?

These questions are certainly not trivial, as it is easy to exhibit counter-examples of "one-variable-at-a-time" methods of optimization that do not converge to the true optimum. This paper answers the above questions in the affirmative for the three following classes of discrete optimal control problems:

(1) States are bounded, controls are unbounded, and the elimination of controls (by virtue of the invertibility of the dependence of states on controls, as described in Sec. IV) results in a convex objective function of states only.

(2) Controls are bounded, states are unbounded, and the elimination of states results in a convex objective function of controls (and initial state).

(3) Both states and controls are bounded, and the objective function is quadratic.

Some related results in successive optimization are found in Ref. 5 for convex programming and in Refs. 6 and 7 for quadratic programming. Our search for a proof of convergence of the DPSA method for the control problem case was motivated by these results. In fact, in each of the three cases treated here, convergence will be proved by relating the control problem to a certain convex or quadratic programming problem.

II FORMULATION OF OPTIMAL CONTROL PROBLEMS SUITABLE FOR DYNAMIC PROGRAMMING

The formulation of optimal control problems assumed here (with slight changes in notation) is somewhat less general than in Ref. 3. The objective function to be optimized is a cost (to be minimized)

$$J = \sum_{k=0}^{K} \ell_k(x_k, u_k) \tag{1}$$

where the minimization is over all trajectory and control sequences, $\{x_k\}_{k=0}^{K}$, $\{u_k\}_{k=0}^{K-1}$ satisfying certain constraints:

(1) The system equations (nonlinear, in general)

$$x_{k+1} = g(x_k, u_k) \quad , \quad k = 0, \ldots, K-1 \tag{2}$$

(2) Control bounds*

$$\underline{u}_k^i \le u_k^i \le \bar{u}_k^i \quad , \quad i = 1, \ldots, n \quad k = 0, \ldots, K-1 \tag{3}$$

---

* The methods of proof here do not appear to be extendible to more general constraints on states and controls as in Ref. 3.

(3) State bounds†

$$\underline{x}_k^i \le x_k^i \le \bar{x}_k^i \quad , \quad i = 1, \ldots, n \quad k = 0, \ldots, K \quad . \tag{4}$$

III DETAILS OF THE DPSA TECHNIQUE

In order to prove convergence of this technique, we must review in detail the way it functions. The technique proceeds as follows:

(1) The states x_k^i are quantized, resulting in a finite set of allowable values for each state. The controls are not quantized to avoid the need for interpolation in the dynamic programs.

(2) Initial <u>feasible</u> control and trajectory sequences

$$\{u_k^{(0)}\}_{k=0}^{K-1} \quad , \quad \{x_k^{(0)}\}_{k=0}^{K}$$

are selected, such that they satisfy Eqs. (2) to (4), with the $x_k^{(0)i}$'s assuming only the quantized values.

(3) One state component index--say, i--is selected. All states x_k^j, $j \neq i$, $k = 0, \ldots, K$ are kept fixed at the values $x_k^{(0)j}$. It is then desired to minimize J over all x_k^i's, $k = 0, \ldots, K$. This 1-dimensional program proceeds in the usual way, except that here there are n controls to vary at each time stage, subject to the $n-1$ states x_k^j, $j \neq 1$ being fixed, thereby leaving one degree of freedom in the controls to allow varying x_k. For this purpose, it must be assumed that the mapping from the controls u_k^j onto the states x_k^j, $j = 1, \ldots, n$ in the control law Eq. (2), is invertible.

(4) When the optimal control and trajectory have been derived using state component x^i, the method proceeds with Step (3) for another state component x^j, until all the components for $i = 1, \ldots, n$ have been treated once.

(5) The method proceeds to repeat Steps (3) and (4) until either insignificantly small reductions in J occur, or no reduction in J occurs at all. At this stage, the process is terminated, and either the last or any recent suitable control sequence is taken as the optimal solution.

IV PROOF OF CONVERGENCE WHEN ONLY STATES ARE BOUNDED

---

†Fixed initial or terminal states are representable by collapsed upper and lower bounds; for free states at any time k, we allow infinite bounds here. Note: this is slightly more general than in Ref. 3, where only a fixed initial state was considered.

A. The Class of Problems

The proof of convergence for a restricted class of problems of the type given by Eqs. (1) to (4) will now be obtained by showing that the DPSA method lowers the cost J at least as fast as a certain convex programming procedure that is known to be convergent.

Class I of such problems shall be defined by the following properties (which are related to the assumptions made in Ref. 5):

I.1 The vector functions $g_k(x_k, u_k)$ in Eq. (2) satisfy an "implicit function theorem" with respect to u_k--i.e., there are (vector) functions $h_k(x_k, v_k)$ such that if for some v,

$$v = g_k(x_k, u_k)$$

then

$$u_k = h_k(x_k, v)$$

i.e., h_k satisfies

$$g_k[x_k, h_k(x_k, v)] \equiv v \qquad \forall\ x_k, k \ .$$

This is the "invertibility" property alluded to in Secs. I and III.

I.2 The cost function obtained by substitution of x_k's for u_k's,

$$\tilde{J} = \sum_{k=0}^{K-1} q_k(x_k, x_{k+1}) + \ell_K(x_K) \tag{5}$$

where

$$q_k(x_k, x_{k+1}) = \ell_k[x_k, h_k(x_k, x_{k+1})] \ ,$$

is convex (although not necessarily strictly convex) and attains a unique minimum for x_k^i between its bounds given by Eq. (4) (if it has any) when all other x_ℓ^j's are fixed. (This is D'Esopo's definition of "convex '" in Ref. 5). Also, $\tilde{J}$ has continuous partial derivatives with respect to each x_k^i.

I.3 Either all the state bounds in Eq. (4) are finite, or for each finite number r the set of states x satisfying Eq. (4) and $\tilde{J}(x) \le r$ is bounded.

A particular subclass of Class I problems of practical interest is the case where

$$\ell_k(x_k, u_k) = \tfrac{1}{2}[x_k' Q_k x_k + u_k' R_k u_k] + A_k x_k + B_k u_k \ , \tag{6}$$

$$k = 0, \ldots, K-1$$

$$\ell_K(x_K) = \tfrac{1}{2} x_K' Q_K x_K + A_K x_K \tag{7}$$

with the Q_k's all positive semi-definite, except that Q_0 is positive definite (unless the initial state is fixed, in which case

there is no Q_0), the R_k's all positive definite, and

$$g_k(x_k, u_k) = \Phi_k x_k + \Gamma_k u_k \quad , \quad k = 0, \ldots, K-1 \tag{8}$$

with the Γ_k's all invertible. It is easily verified that Eqs. (6) to (8) imply properties I.1, I.2, and I.3.

B. The Successive Optimization Method

Consider the following method (which we shall refer to as the SO method) for solving the problem of minimizing the cost $\tilde{J}$ in Eq. (5) subject Eq. (4):

(1) Pick an initial point $x^{(0)}$ satisfying Eq. (4).

(2) Start with the indices $i = 1$ and $k = K$ for x_k^i.

(3) Minimize $\tilde{J}(x)$ over all quantized levels for x_k^i, subject to all other variables x_ℓ^j fixed.

(4) Define $x^{(1)}$ by using the newly obtained value of x_k^i (which may equal the previous value) and the other $x_\ell^{(0)j}$'s.

(5) Repeat (2) for k decreased by 1, unless k is 1, in which case go on to Step (3) with the next higher value of i, and set $k = K$, unless $i = n$, in which case go back to Step (2).

(6) Keep repeating the above Steps (2) to (5) until either (a) a vector $x = (x_k^i \mid k = 0, \ldots, K, \; i = 1, \ldots, n)$ is reached for which the cost $\tilde{J}$ cannot be reduced at all, or (b) the cost $\tilde{J}[x^{(N)}]$ is converging to a finite limiting value.

Lemma 1: If the DPSA and SO methods begin with the same initial trajectory sequence $\{x_k^{(0)i}\}$, then after N 1-dimensional dynamic programs the DPSA method for problem (1) to (4) attains a value $J[x^{(N)}, u^{(N)}]$ at least as low as the value $\tilde{J}[x^{(NK)}]$ of the NK^{th} iteration of the SO method for the problem defined by Eqs. (4) and (5).

Proof: This lemma follows trivially from the equivalence of problems defined by Eqs. (1) to (4) and the problem defined by Eqs. (4) and (5), and the fact that minimizing $J(x, u)$ over all x_k^i's, $k = 0, \ldots, K$ (for a fixed i) can certainly do no worse than minimizing $J(x, u)$ over all x_k^i's one at a time as k goes from K to 0 (assuming the 1-dimensional dynamic programs in the DPSA method are the "backward" type).

Lemma 2: For continuous states x_k^i the SO method for problem defined by Eqs. (4) and (5) converges to a true optimum solution under assumptions I.1 to I.3--i.e., the costs $\tilde{J}[x^{(N)}]$ converge to the optimal cost

$\tilde{J}^{(\infty)}$, and there is at least one limit point $x^{(\infty)}$ of the sequence $\{x^{(N)}\}_{N=1}^{\infty}$, with $\tilde{J}[x^{(\infty)}] = \tilde{J}^{\infty}$. Furthermore, if $\tilde{J}(x)$ is strictly convex in x, then there is only one limit point $x^{(\infty)}$, and this is the unique optimal solution.

Proof: This result is a restatement of some of the conclusions in the Theorem of Ref. 5, whose assumptions include properties I.1 to I.3 as a special case.

Lemma 3: If the quantization levels in problem defined by Eqs. (4) and (5) are suitably arranged and sufficiently small, the SO method will converge to a quantized vector $x^{(\infty)}$ having the lowest possible cost, in a finite number of iterations, and this cost can be brought as close as desired to the optimal cost of the unquantized problem.

Proof: Given our assumptions I.1 to I.3, we know by Lemma 2 that the method converges to an optimal solution when the unique true optimum is obtained at each iteration with respect to one variable at a time. When the states x_k^i are quantized, it may happen that the optimum with respect to x_k^i occurs at two quantized values. Under continuous variation of x_k^i, our assumption I.2 requires that there be only one minimum. Thus we can simply assume that the quantized level has been perturbed ever so slightly, such that there is only one minimum here without having disturbed the minimality at any previous iteration. Now, the continuous problem has an optimum; hence, the cost in the quantized case is bounded below. But there is only a finite number of possible quantized state-level combinations for all the x_k^i's. Since, at any given iteration, the above level perturbations produce a strictly decreasing cost, or else no change in x_k^i occurs at that iteration, it follows that either the optimal quantized cost is attained in a finite number of iterations, or the iterations stop changing the x_k^i's after some finite number of iterations, and the quantized cost attained is not optimal among the quantized costs obtainable. (The latter situation can occur--for example, as in Fig. 1--where the continuous problem's convex cost function has a very narrow valley relative to the quantization levels, and the SO method for the quantized problem can in fact stop at a non-optimal point). However, in the latter case one can merely increase the number of quantized states until a definite change in state is forced at each iteration, up to any desired number of iterations. Consequently, we conclude that if the quantized states are made sufficiently fine, the method must attain the lowest (quantized) cost, and of course, as the quantization is made smaller and

smaller, due to continuity of the convex function $\tilde{J}(x)$, the quantized optimum must converge to the true optimum.

Combining Lemmas 1 and 3, we obtain the following conclusion regarding the DPSA method for problem defined by Eqs. (1) to (4):

Theorem: Suppose that an optimal control problem given by Eqs. (1), (2), and (4) satisfies properties I.1 to I.3. Then, for a suitable quantization of state variables, the DPSA method will converge to an optimal solution of the quantized problem. As the quantization increments approach zero, the quantized problem's solutions converge to an optimal solution of the continuous problem.

V PROOF OF CONVERGENCE WHEN ONLY CONTROLS ARE BOUNDED

In the case where there are no bounds on states, except possibly on x_0, a result similar to that in Sec. IV is obtainable by elimination of the states x_k^i, $k = 1, \ldots, K$, $i = 1, \ldots, n$ from the set of optimization variables, thereby leaving $x_0, u_1, \ldots, u_{K-1}$ as independent vectors subject only to Eq. (3) and the part of Eq. (4) pertaining to x_0.

Let Class II be the subclass of optimal control problems, characterized by Eqs. (1) to (4), for which the bounds in Eq. (4) do not exist (i.e., may be regarded as infinite), except possibly the ones on the initial state x_0, and having the following properties:

II.1 Same as I.1 (this is needed to be able to apply the DPSA method).

II.2 The cost function $\tilde{J}(x_0, u_1, \ldots, u_{K-1})$, obtained by eliminating the variables x_k, $k = 1, \ldots, K$, by means of Eq. (2) in terms of x_0 and the u_k's, $k = 0, \ldots, K-1$, is convex and possesses a unique minimum with respect to each x_0^i, u_k^i, $i = 1, \ldots, n$, $k = 0, \ldots, K-1$.

II.3 The control bounds, Eq. (2), and bounds on x_0 are all finite, or at least the set of $(x_0, u_1, \ldots, u_{K-1})$ satisfying Eq. (3) and the bounds on x_0, for which $\tilde{J}(x_0, u_1, \ldots, u_{K-1}) \leq r$, is bounded for every finite number r.

It can easily be checked that linear control laws Eq. (8) with quadratic costs Eq. (6) and (7), in which the Γ_k's are invertible and Q_0 and the R_k are positive definite, while the remaining Q_i's are positive semidefinite, do in fact satisfy II.1 to II.3. (Note: This is the same class of control problems that was said to satisfy I.1 to I.3, except that now the states are unbounded instead of the controls.)

Here, we may use the same result[5] as in Sec. IV to obtain exactly the same conclusion as in the theorem of Sec. IV, with conditions I.1 to I.3

replaced by conditions II.1 to II.3.

VI PROOF OF CONVERGENCE IN THE GENERAL CASE WITH BOTH STATES AND CONTROLS BOUNDED

When both states and controls are bounded by constraints of the form given by Eqs. (3) and (4), it is not possible to eliminate controls or states to derive an equivalent optimization problem without equality constraints for which convergence is known. The reason is that any attempt to eliminate equality constraints would lead to additional inequality constraints on the remaining variables, which would not be of the simple form given by Eq. (3) or (4), and convergence proofs for one-variable-at-a-time methods are known only for these simple types of constraints.[5-7] In fact, it is very easy to exhibit counterexamples where one-at-a-time methods fail for general constraints.

Thus we shall take another approach, by generalizing another of the known proofs of convergence,[6] to the case where equality constraints exist. In this case the optimal control problem will need to be more restricted.

Let Class III of the optimal control problems described by Eqs. (1) to (4) be characterized by the following properties:

III.1 The control law Eq. (2) is linear, as in Eq. (8), with the Γ_k's all invertible. (Note: This fact will be needed in the proof to follow and not just to guarantee the DPSA method to work, as was the case in the two earlier sections.)

III.2 The cost function Eq. (1) is as given by Eqs. (5) and (6), with the matrices Q_k, $k = 0, \ldots, K$, and R_k, $k = 0, \ldots, K-1$, all positive definite.

III.3 Either the bounds Eqs. (3) and (4) are all finite, or the set $\{(x, u) \mid J(x, u) \leq r, \ (x, u)$ satisfies Eqs. (3) and (4)$\}$ is bounded for each finite number r.

For Class III we obtain exactly the same conclusion as in the theorem of Sec. IV, by the same approach as in Sec. IV. In this case, however, we cannot rely upon a known convergence proof for quadratic programming, as we did for the convex programming situation in earlier sections, since there appear to be no proofs of convergence for successive optimization methods in quadratic programming with general linear constraints. The results in Refs. 6 and 7 do consider general constraints of the form

$$Ax \geq b$$

or

$$Ax = b, \quad x \geq 0,$$

but the proofs of convergence given there (incidentally, the proof in Ref. 7

is incorrect for the semidefinite case) are for the dual quadratic program (i.e., one solves the dual problem by a convergent method and then uses the answer to derive a solution for the primal problem). The dual problem in Refs. 6 and 7, however, has only the simple constraint $x \geq 0$; consequently the result is of no use here. The idea introduced there, of passing to the dual problem, is a very interesting one and we shall use it to prove convergence in the present case.

Lemma 4: Suppose that $\{x_k\}_{k=0}^{K}$ and $\{u_k\}_{k=0}^{K-1}$ are a trajectory and control for a problem in Class III for which the DPSA method cannot decrease the cost any further (assuming no quantization--i.e., continuous variation of the state variables). Then the method has attained the true (unique) optimum for the given control problem.

Proof: We can assume (without loss of generality) that the linear terms in Eqs. (6) and (7) have been eliminated by suitable choice of coordinates in control and state space. Let Lagrange multiplier vectors λ_k, $k = 0, \ldots, K-1$ be associated with the Eq. (2), which are Eq. (8) in this case, and ν_k, μ_k with the inequalities given by Eqs. (3) and (4) respectively. Interpret ν_k^i as associated with the upper bound $\bar{u}_k^i$ when it is positive and the lower bound when it is negative (similarly for μ_k^i). Writing the Lagrangian in the form

$$\mathcal{L} = J + \sum_{k=0}^{K-1} \lambda_k (x_{k+1} - \Phi_k x_k - \Gamma_k u_k)$$

$$+ \sum_{k=0}^{K} [\bar{\mu}_k (x_k - \bar{x}_k) - \underline{\mu}_k (x_k - \underline{x}_k)]$$

$$+ \sum_{k=0}^{K-1} [\bar{\nu}_k (u_k - \bar{u}_k) - \underline{\nu}_k (u_k - \underline{u}_k)] \quad ,$$

and using the Kuhn-Tucker Theorem (which yields necessary and sufficient conditions for an optimum in the case of linear constraints and a convex quadratic objective function), we obtain the following criteria for optimality:

The sequences of vectors $\{x_k\}_{k=0}^{K}$ and $\{u_k\}_{k=0}^{K-1}$ are the optimal trajectory and control if and only if there exist λ_k, μ_k, and ν_k, such that

$$\Gamma_k^T \lambda_k = R_k u_k + \nu_k \quad , \qquad k = 0, \ldots, K-1 \tag{9}$$

$$\Phi_k^T \lambda_k = \lambda_{k-1} + Q_k x_k + \mu_k \quad , \quad k = 0, \ldots, K \tag{10}$$

$$\mu_k^i \begin{cases} \geq 0 & \text{if } x_k^i = \bar{x}_k^i \\ \leq 0 & \text{if } x_k^i = \underline{x}_k^i \\ = 0 & \text{otherwise} \end{cases} \quad , \quad i = 1, \ldots, n \quad , \quad k = 0, \ldots, K-1 \tag{11}$$

$$\nu_k^i \begin{cases} \geq 0 & \text{if } u_k^i = \bar{u}_k^i \\ \leq 0 & \text{if } u_k^i = \underline{u}_k^i \\ = 0 & \text{otherwise} \end{cases} \quad , \quad i = 1, \ldots, n \quad , \quad k = 0, \ldots, K-1 \tag{12}$$

The additional vectors λ_{-1}, λ_K introduced in Eq. (10) are taken to be the 0 vector.

We shall now prove optimality for the given trajectory and control reached by the DPSA method by showing that Lagrange multipliers satisfying Eqs. (9) to (11) exist. For termination of the DPSA method, we know that the given control and trajectory provide an optimum for the restricted problem where only the x_k^i's are varied for some i while all other x_k^j's are fixed. For each fixed i the Kuhn-Tucker conditions for the optimum (which also are necessary and sufficient for the same reason) are precisely the same, except that Eqs. (10) and (11) hold only for the i^{th} component--i.e., there are Lagrange multiplier vectors $\lambda_k^{(i)}$, $\nu_k^{(i)}$, and scalars μ_k^i satisfying Eqs. (9), (12)

$$(\Phi_k^i)^T \lambda_k^{(i)} = \lambda_{k-1}^{(i)i} + Q_k^i x_k + \mu_k^i \quad , \quad k = 0, \ldots, K$$

and Eq. (11i), which is (11) for that i only (where A^i denotes the i^{th} row of matrix A), and Eq. (10i). We shall show that the vectors $\lambda_k^{(i)}$ are in fact the same for all i, from which it will immediately follow from Eq. (9) that the $\nu_k^{(i)}$'s are independent of i, and from Eq. (10) that the μ_k^i's comprise the components of appropriate vectors μ_k, which together satisfy Eqs. (9), (10), (11), and (12), thereby proving optimality for the global optimal control problem.

We first derive the uniqueness of the λ_k's for the global program by the following:

Lemma 5: The Lagrange multipliers for a quadratic optimal control problem with linear constraints satisfying conditions III.1 to III.3 are unique.

Remark: Actually we will not need this result, but an analogous result for the $\lambda_k^{(i)}$'s will follow Lemma 5.

Proof: The ideas in this proof come from Ref. 6.

Combining Eqs. (9), (10) with Eqs. (2), (8), we may eliminate the

x_k's and u_k's to obtain the relations

$$Q_{k+1}^{-1}\,\Phi_{k+1}^{T}\,\lambda_{k+1} + \Phi_k\,Q_k^{-1}\,\lambda_{k-1}$$
$$- [Q_{k+1}^{-1} + \Phi_k\,Q_k^{-1}\,\Phi_k^{T} + \Gamma_k\,R_k^{-1}\,\Gamma_k^{T}]\,\lambda_k$$
$$- Q_{k+1}^{-1}\,\mu_{k+1} + \Phi_k\,Q_k^{-1}\,\mu_k + \Gamma_k\,R_k^{-1}\,\nu_k = 0$$
$$k = 0, \ldots, K-1 \qquad (13)$$

for the Lagrange multipliers alone. But Eqs. (11), (12), and (13) may be interpreted (by inspection) as the Kuhn-Tucker conditions for the quadratic programming problem (called the <u>dual</u> problem):

$$\min_{\lambda_k}\ \tilde{J}(\lambda) = \min_{\substack{k=0,K\\ k=0}}^{K-1} \sum \left[\frac{1}{2}\,\lambda_k^{T}\left\{Q_{k+1}^{-1} + \Phi_k\,Q_k^{-1}\,\Phi_k^{T} + \Gamma_k\,R_k^{-1}\,\Gamma_k^{T}\right\}\lambda_k - \lambda_k^{T}\,Q_{k+1}^{-1}\,\Phi_{k+1}^{T}\,\lambda_{k+1}\right] \qquad (14)$$

subject to the constraints

$$Q_k^{-1}\,\Phi_k^{T}\,\lambda_k - Q_k^{-1}\,\lambda_{k-1} \begin{cases} \geq \underline{x}_k \\ \leq \bar{x}_k \end{cases} \qquad k = 0, \ldots, K \qquad (15)$$

$$R_k^{-1}\,\Gamma_k^{T}\,\lambda_k \begin{cases} \geq \underline{u}_k \\ \leq \bar{u}_k \end{cases} \qquad k = 0, \ldots, K-1 \qquad (16)$$

where μ_k's and ν_k's are the Lagrangian vectors associated with Eqs. (15) and (16), respectively. [To check this, note that the first three terms in Eq. (13) arise from differentiating Eq. (14) with respect to λ_k, and the remaining terms by differentiating the terms in the Lagrangian for Eqs. (15) and (16).] Now, it is easy to determine whether the matrices

$Q_{k+1}^{-1} + \Phi_k\,Q_k^{-1}\,\Phi_k^{T} + \Gamma_k\,R_k^{-1}\,\Gamma_k^{T}$	$-Q_{k+1}^{-1}\,\Phi_{k+1}^{T}$
$-\Phi_{k+1}\,Q_{k+1}^{-1}$	$Q_{k+2}^{-1} + \Phi_{k+1}\,Q_{k+1}^{T}\,\Phi_{k+1}^{T} + \Gamma_{k+1}\,R_{k+1}^{-1}\,\Gamma_{k+1}^{T}$

are positive definite. But these appear as the largest principal submatrices for the large matrix obtained when the λ_k's are arranged in sequence in writing $\tilde{J}$ as a quadratic form of the type

$$\lambda^{T}\,A\,\lambda\ ,$$

and it is easy to check that positive definiteness of these assures positive definiteness of $\tilde{J}$ as a function of variables λ_k^i. Thus $\tilde{J}$, being a strictly convex function, attains a unique

minimum over the convex set determined by Eqs. (15) and (16). This proves Lemma 5 for the global problem.

The same conclusion holds for each subproblem of the DPSA method, where all x_k^j's are held fixed, $j \neq i$, for some i. This is seen as follows: If we have Eqs. (10i) and (11) for i only, we can still use Eq. (10) to eliminate the x_k's by interpreting x_k as

$$x_k = Q_k^{-1}\left[\Phi_k^T \lambda_{k-1}^{(i)} - \mu_k^{(i)}\right] ,$$

where $\mu_k^{(i)}$ is defined by Eq. (10), and therefore has its i^{th} component equal to μ_k^i. This leads to Eq. (13), just as before, except that now $\mu_k^{(i)j}$, $j \neq i$, no longer has necessarily the right sign according to Eq. (11). Thus we obtain the interpretation Eqs. (14), (15i), and (16), where Eq. (15i) is Eq. (15), with the inequalities reversed wherever necessary so that $\mu_k^{(i)j}$ has the right sign, $j \neq i$. This results in a quadratic program which is dual to the original problem modified in the direction of its inequalities Eq. (4) (with respect to the attained values of x_k^j, $j \neq i$, instead of the actual constraints $\bar{x}_k^j$, $\underline{x}_k^j$). Nevertheless, the same conclusion results--i.e., that a unique optimum for $\tilde{J}$ is attained. This proves the uniqueness of the $\lambda_k^{(i)}$'s.

The proof of Lemma 4 now follows easily: If for some i, $\mu_k^{(i)j}$ [defined by (10) as previously discussed] has the right signs for $j \neq i$, $i = 1, 2, \ldots, n$, according to (11), then the $\lambda_k^{(i)}$'s, $\nu_k^{(i)}$'s, and $\mu_k^{(i)}$'s would constitute a set of multipliers for the global problem. Thus, suppose no $\mu_k^{(i)}$'s have the right signs in all their components for any fixed i. If the $\lambda_k^{(i)}$'s were independent of i, then by virtue of Eq. (10i), relations (10) and (11) would be satisfied by the vectors μ_k comprised by the components $\mu_k^i = \mu_k^{(i)i}$, and again we would have multipliers for the global problem. Thus the only remaining possibility is that there are distinct vectors $\lambda_k^{(i)}$ (for some k), and the corresponding $\mu_k^{(i)}$'s do not have the right signs in all of their components. We now obtain a contradiction. Taking linear combinations $\tilde{\lambda}_k = \alpha_i \lambda_k^{(i)} + \alpha_j \lambda_k^{(j)}$ of two distinct $\lambda_k^{(i)}$'s with $\alpha_i + \alpha_j = 1$, $\alpha_i, \alpha_j > 0$ certainly yields a $\tilde{\nu}_k$ by Eq. (9) that satisfies Eq. (12). Now if $\mu_k^{(i)j}$ and/or $\mu_k^{(j)i}$ had the wrong sign, then the vector $\tilde{\mu}_k$ defined by (10) using $\lambda_k^{(i)}$ and $\lambda_{k-1}^{(j)}$, being linear in $\mu_k^{(i)}$ and $\mu_k^{(j)}$, would still have the right sign in the component $\tilde{\mu}_k^i$ for sufficiently small $\alpha_j > 0$, thereby contradicting the uniqueness of the Lagrange multipliers $\lambda_k^{(i)}$, $\lambda_k^{(j)}$ for their

respective subproblems in the DPSA method. Thus, in fact, both $\mu_k^{(i)j}$ and $\mu_k^{(j)i}$ must have been zero. Since this is true for any i, j, k, the final conclusion is that the $\lambda_k^{(i)}$'s and the ν's and μ's comprise a set of multipliers for the global problem; hence, $\{x_k\}_{k=0}^{K}$ and $\{\mu_k\}_{k=0}^{K-1}$ are optimal. This completes the proof of Lemma 4.

Using Lemma 4, precisely the same convergence conclusions as in Secs. IV and V result for the DPSA method applied to problems in Class III.

VII APPLICATIONS

One area to which extensive applications of the DPSA technique have been made is the optimization of multipurpose water reservoir systems. A typical problem, which is discussed in Larson and Keckler,[8] is shown in Fig. 2. The four-reservoir system is to be operated over a 24-hour period such that benefits from power generation and irrigation are maximized, subject to constraints for flood control and recreational use. The system equations assume a transport delay between releases at an upstream reservoir and inflow into a downstream reservoir. For this problem the system equations, performance criteria, and constraints were linear. Convergence to the true optimum was obtained in 30 seconds on the B-5500 computer.

Optimum yearly operation of the six reservoirs on the Missouri River main stem has been studied by Larson.[4] Actual data was used in this application.[9] Other studies of optimum reservoir operation have been made by Fukao and Nureki,[10] Bernholtz et al.,[11] and Anstine and Ringlee.[12]

A related area is the optimal planning of additions to power transmission subsystems. This application is considered by Peschon, Kaltenbach, et al.[13]

A modified version of the DPSA technique was used in an airline scheduling algorithm developed by Larson.[3] The objective in this case is to operate a given fleet of aircraft over a fixed system of routes so as to maximize profit. One state variable is required for each aircraft. For the airline system shown in Fig. 3, solutions have been obtained for problems involving as many as 100 aircraft.

Another area where the method can be used is in the computation of the discrete-time Ricatti equations.[14] These equations occur in the solution of optimization problems with linear system equations, a quadratic performance criterion, and no constraints; thus, the DPSA technique is guaranteed to converge (see Secs. IV to VI). For the case where $m = n$, the n-dimensional Ricatti equation, which requires inversion on the $n \times n$ matrix at each time instant, can be broken down into a sequence of one-dimensional recursive relations. The optimal control and minimum cost function take the form

$$\hat{u}_k^i = A_k^i x_k^i + a_k^i$$

$$J_k = B_k^i (x_k^i)^2 + b_k^i x_k^i + c_k^i$$

where scalar recursive equations can be written for A_k^i, a_k^i, B_k^i, b_k^i, and c_k^i. The computational advantages of this approach in high-dimensional linear control and estimations problems[14-15] can be very great indeed.

VIII CONCLUSIONS

This paper has presented proofs that the DPSA technique converges to the true optimum solution in three important classes of problems. The conditions under which the proofs are valid provide guidelines for determining whether or not convergence will be obtained in a particular example. Thus, applications in the areas discussed in Sec. VII as well as entirely new problems can be attacked by this method on a mathematically sound basis.

REFERENCES

1. R. E. Bellman, Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1957.

2. R. E. Bellman, S. E. Dreyfus, Applied Dynamic Programming, Princeton University Press, Princeton, Jew Jersey, 1962.

3. R. E. Larson, State Increment Dynamic Programming, American Elsevier Press, New York, New York, 1967.

4. R. E. Larson, "A Dynamic Programming Successive Approximations Technique," Proc. Joint Automatic Control Corference, University of Michigan, June 1968.

5. D. A. D'Esopo, "A Convex Programming Procedure," Naval Research Logistics Quarterly, Vol. 6, pp. 33-42 (1959).

6. H. P. Künzi, W. Krelle, W. Oettli, (Translated by F. Levin), Nonlinear Programming, Blaisdell, Waltham, 1966.

7. C. Hildreth, "A Quadratic Programming Procedure," Naval Research Logistics Quarterly, Vol. 4, pp. 79-85 (1957).

8. R. E. Larson, W. G. Keckler, "Applications of Dynamic Programming to the Control of Water Resource Systems," Proc. IFAC Symposium on Computer Control of Natural Resources and Public Services, Haifa, Israel, September 1967.

9. "Use of Computer Techniques for Regulation of Missouri River Main Stem Reservoir System," U.S. Army Engineer Division, Missouri River, Omaha, Nebraska (January 1966).

10. T. Fukao, R. Nureki, "Successive Approximations Methods of Dynamic Programming in Reservoir Control Problems," Bull. Electro-Tech Lab., Vol. 26, No. 3, pp. 178-202, Tokyo, 1962.

11. B. Bernholtz, L. J. Graham, "Hydrothermal Economic Scheduling, Part I: Solution by Incremental Dynamic Programming," AIEE Trans. on Power Apparatus and Systems (Part III), December 1960, pp. 921-932; also "Part II: Extension of the Basic Theory," ibid., February 1962, pp. 1089-1096.

12. L. T. Anstine, R. J. Ringlee, "Susquehanna River Short-Range Hydro-Thermal Coordination," IEEE Trans. on Power Apparatus and Systems, pp. 185-190, April 1963.

13. J. Peschon, J. C. Kaltenbach, P. Henault, M. W. Siddiqee, L. P. Hajdu, "Power System Planning and Reliability," prepared at Stanford Research Institute, Menlo Park, California, for Bonneville Power Administration, Portland, Oregon, June 1968.

14. R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," Trans. ASME, J. Basic Engr., Vol. 83, Series D, No. 1, pp. 95-108 (March 1961).

15. R. E. Kalman and R. W. Koepcke, "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes," Trans. ASME, Vol. 80, No. 8, pp. 1820-1838 (November 1958).

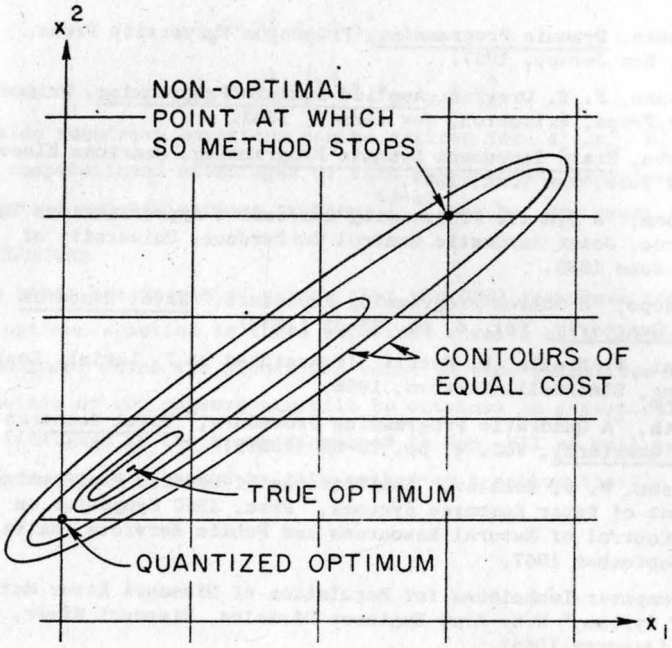


Fig. 1 Effect of Quantization on the SO Method for Two Dimensions

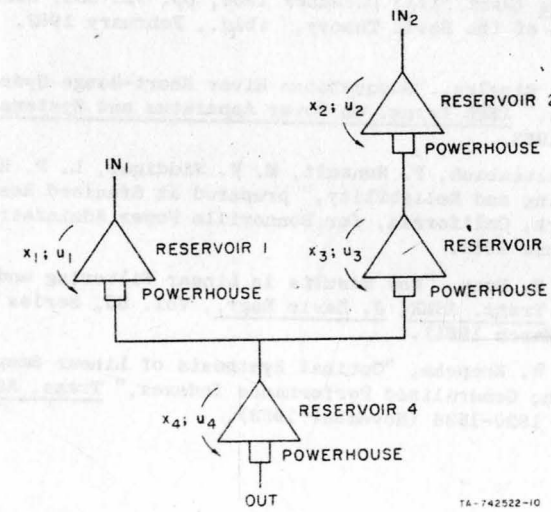
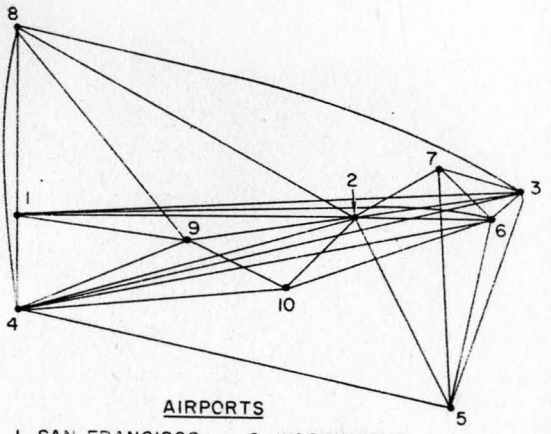


Fig. 2 Reservoir Configuration



AIRPORTS

1 SAN FRANCISCO	6 WASHINGTON
2 CHICAGO	7 DETROIT
3 NEW YORK	8 SEATTLE
4 LOS ANGLES	9 DENVER
5 MIAMI	10 KANSAS CITY

Fig. 3 Airline System Map