Read

The Conjugate Gradient Method for Optimal Control Problems

L. S. LASDON, MEMBER, IEEE, S. K. MITTER, AND A. D. WAREN, MEMBER, IEEE

Abstract—This paper extends the conjugate gradient minimization method of Fletcher and Reeves to optimal control problems. The technique is directly applicable only to unconstrained problems; if terminal conditions and inequality constraints are present, the problem must be converted to an unconstrained form; e.g., by penalty functions. Only the gradient trajectory, its norm, and one additional trajectory, the actual direction of search, need be stored. These search directions are generated from past and present values of the objective and its gradient. Successive points are determined by linear minimization down these directions, which are always directions of descent. Thus, the method tends to converge, even from poor approximations to the minimum.

Since, near its minimum, a general nonlinear problem can be approximated by one with a linear system and quadratic objective, the rate of convergence is studied by considering this case. Here, the directions of search are conjugate and hence the objective is minimized over an expanding sequence of sets. Also, the distance from the current point to the miminum is reduced at each step.

Three examples are presented to compare the method with the method of steepest descent. Convergence of the proposed method is much more rapid in all cases. A comparison with a second variational technique is also given in Example 3.

I. Introduction

THIS PAPER presents an iterative procedure for solving unconstrained optimal control problems. Of course, a general formulation of the optimal control problem involves both terminal constraints on the state variables and inequality constraints on the state and control variables enforced along the entire trajectory. Penalty functions have often been used to convert such problems to a sequence of "unconstrained" problems, i.e., problems with no terminal or inequality constraints [1]–[4]. It is evident that the efficiency of these methods depends greatly on the technique used to solve the unconstrained optimal control problem. Presently available techniques all have shortcomings. The convergence of steepest descent methods is often slow [1] whereas second-variational and Newton methods may not converge at all. Thus there is strong motivation for developing more efficient means for solving unconstrained optimal control problems.

Manuscript received August 11, 1966; revised December 30, 1966. This work was supported in part by U. S. Army Research Office Grant DA-AROD-31-124-6647 and by National Science Foundation Grant GK-600.

L. S. Lasdon is with the Division of Organizational Sciences, Operations Research Group and Systems Research Center, Case Institute of Technology, Cleveland, Ohio.

S. K. Mitter is with the Engineering Division and Systems Research Center, Case Institute of Technology, Cleveland, Ohio.

A. D. Waren is with the Dept. of Electrical Engineering, Fenn College of Engineering, Cleveland State University, Cleveland, Ohio.

Similar difficulties existed, until recently, in the field of finite dimensional optimization, i.e., mathematical programming. However, in the past few years several rapidly convergent finite dimensional unconstrained minimization techniques have been developed. Among these are the method of Fletcher and Powell [5] and the Fletcher–Reeves [6] adaptation of the conjugate gradient method of Hestenes and Stiefel [7]. Both these procedures generate conjugate directions of search and therefore minimize a positive definite quadratic function of n variables in n steps. In addition, the directions generated are always directions of descent and thus, for relatively smooth functions, the function value is decreased at each step.

The combination of these properties implies that the methods converge rapidly to the nearest local minimum for a general function of n variables. Experience has shown that both techniques converge much more rapidly, in general, than the method of steepest descent while requiring only function and gradient evaluations. Moreover, their stability properties are superior to those of second-order Newton methods, and second-order derivatives are not required.

Function space analogs of the steepest descent and second-order Newton techniques have been developed and applied to problems of optimal control. In particular, Kelly and Bryson [8], [9], Mitter [10], and others have developed steepest descent and second-order methods. However, the analogs of the conjugate direction techniques have not yet been forthcoming. Since these methods are considered by some authors [5], [11] to be the most powerful presently available for finite dimensional minimization problems, it seems appropriate to consider their generalization to optimal control.

This paper describes an extension of the Fletcher–Reeves [6] conjugate gradient method to function space problems. The computational simplicity of this algorithm led to its selection. As in the steepest descent method, the gradient trajectory must be computed and stored. In addition, the conjugate gradient technique requires the computation of the norm of the gradient and the storage of one other trajectory, the actual direction of search.

Despite its simplicity, computational results illustrate its marked superiority to the method of steepest descent. These results are substantiated by theoretical developments.

It is shown here that the directions in function space generated by the conjugate gradient method are such that the objective function is decreased at each step. This leads to the results shown in Section III concerning convergence from arbitrary starting points as the number of iterations approaches infinity. The rate of convergence is best investigated by considering the performance of the algorithm on a problem with a linear system and quadratic objective (Section IV). Here the results of Antosiewicz and Rheinboldt [12] apply, showing that the directions of search are conjugate and hence the function is minimized over an expanding sequence of sets. In addition, the distance from the current estimate of the solution to the optimal point decreases monotonically.

Further results in the linear-quadratic case have been obtained. For linear-quadratic problems in one state variable, the conjugate gradient method achieves a value of the objective at least as low as that obtained by the method of steepest descent at each iteration (assuming the same starting point for both). Also, a class of quadratic problems is exhibited for which the conjugate gradient method finds the optimal solution in a finite number of steps.

None of the above properties is shared by the method of steepest descent, and this accounts for the rapid convergence of the conjugate gradient method for quadratic problems. This implies that for those general nonlinear problems which may be approximated by linear-quadratic problems near the optimum, the convergence is rapid.

Like most other iterative procedures, this method cannot distinguish between local and global minima. In general, the best that can be expected is efficient convergence to the bottom of whatever valley it starts in [6]. The usual procedure for problems with local minima is to rerun the method with different starting points.

II. Conjugate Gradient Algorithm

A. Problem Formulation

Consider the following problem:

$$\text{minimize} \quad J = \phi(\mathbf{x}(t_f)) \tag{1}$$

$$\text{subject to} \quad \dot{\mathbf{x}} = \boldsymbol{f}(\mathbf{x}, \boldsymbol{u}, \boldsymbol{t}) \tag{2}$$

$$\mathbf{x}(t_0) = \boldsymbol{c} \tag{3}$$

where $\mathbf{x}$ is an n vector, $\boldsymbol{u}$ is an m vector, and $\boldsymbol{t}_0$, $\boldsymbol{t}_f$ are fixed. It is assumed that given a control $\boldsymbol{u}$, (2) and (3) can be solved for a unique $\mathbf{x}=\mathbf{x}(\boldsymbol{u})$, and thus $J=J(\boldsymbol{u})$ is a function of $\boldsymbol{u}$ alone. Furthermore, the existence of the gradient of $J(\boldsymbol{u})$, $\nabla J(\boldsymbol{u})=g(\boldsymbol{u})$ is assumed. The objective function (1) may include penalty function terms to account for constraints.

For the remainder of the paper only the case of a single control function $u(t)$ $(m=1)$ will be considered. The extension to the multicontrol case is straightforward.

B. Algorithm

The conjugate gradient algorithm requires the computation of the gradient trajectory. Let

$$H = \sum_{i=1}^{n} \lambda_i f_i \tag{4}$$

where

$$\dot{\lambda}_i = -\sum_{j=1}^{n} \lambda_j \frac{\partial f_j}{\partial x_i} \tag{5}$$

$$\lambda_i(t_f) = \frac{\partial \phi}{\partial x_i}\bigg|_{t=t_f} \qquad i = 1, \cdots, n. \tag{6}$$

Then the gradient is

$$g(u) = \frac{\partial H}{\partial u}\,. \tag{7}$$

Let $u_i(t)$ be the ith approximation to the optimal control $u^0(t)$. The corresponding gradient $g(u_i)$ is computed by solving the state equations (2) and (3) forwards with $u=u_i$, solving the adjoint equations (5) and (6) backwards and then computing $g(u_i)$ from (7).

The algorithm proceeds as follows:

$$u_0 = \text{arbitrary} \tag{8}$$

$$g_0 = g(u_0) \tag{9}$$

$$s_0 = -g_0. \tag{10}$$

Choose

$$\alpha = \alpha_i \text{ to minimize } J(u_i + \alpha s_i) \tag{11}$$

and then

$$u_{i+1} = u_i + \alpha_i s_i \tag{12}$$

$$g_{i+1} = g(u_{i+1}) \tag{13}$$

$$\beta_i = (g_{i+1}, g_{i+1})/(g_i, g_i) \tag{14}$$

$$s_{i+1} = -g_{i+1} + \beta_i s_i \tag{15}$$

where

$$(g_i, g_j) = \int_{t_0}^{t_f} g_i(t) g_j(t) dt. \tag{16}$$

Note that the new direction of search s_{i+1} is not the negative gradient direction $-g_{i+1}$, but is computed via (15). The distance traveled in this direction is determined by the one-dimensional minimization in (11). Subsequent proofs assume that this minimization is carried out exactly. In practice, this is not possible. Numerical experience has shown that stepping down the search direction until the objective starts to increase in value and then using cubic interpolation gives reasonable results.

III. Convergence

Let the control u be an element of a Hilbert space H and $J(u)$ a Frechet differentiable mapping from H to the real numbers. The conjugate gradient method when applied to $J(u)$ generates directions s_i which are always directions of descent, i.e.,

$$\frac{d}{d\alpha} J(u_i + \alpha s_i)_{\alpha=0} < 0 \tag{17}$$

and this assures that $J(u)$ is decreased at each step. These statements are proved below.

Theorem 1

If $g(u_i) = g_i \neq 0$ then

$$(s_i, g_{i+1}) = 0$$

and

$$\frac{d}{d\alpha} J(u_i + \alpha s_i)\big|_{\alpha=0} = (s_i, g_i) = -\|g_i\|^2.$$

Proof: Let $\alpha = \alpha_i$ minimize $J(u_i + \alpha s_i)$. Then

$$\frac{d}{d\alpha} J(u_i + \alpha s_i)\big|_{\alpha=\alpha_i} = (s_i, g_{i+1}) = 0. \tag{18}$$

From (15)

$$(g_i, s_i) = (g_i, -g_i + \beta_{i-1} s_{i-1}) \tag{19}$$

$$= -\|g_i\|^2 + \beta_{i-1}(g_i, s_{i-1}) = -\|g_i\|^2. \tag{20}$$

Theorem 2

If $g_i \neq 0$ then $J(u_{i+1}) < J(u_i)$.

Proof: Assume there exists no $\alpha > 0$ such that

$$J(u_i + \alpha s_i) < J(u_i). \tag{21}$$

Thus for all $\alpha > 0$

$$\frac{J(u_i + \alpha s_i) - J(u_i)}{\alpha} \geq 0. \tag{22}$$

In the limit as α approaches zero this gives

$$(g_i, s_i) = -\|g_i\|^2 \geq 0 \tag{23}$$

which contradicts the original assumption and hence there exists $\alpha > 0$ such that $J(u_i + \alpha s_i) < J(u_i)$. Since α_i is chosen to minimize $J(u_i + \alpha s_i)$, the theorem is proved.

The sequence of real numbers $\{J(u_i)\}$ is thus monotone decreasing and therefore has a limit J_∞ in the extended real numbers.

Also of interest is the limiting behavior of the sequences $\{u_k\}$ and $\{g_k\}$. Results similar to those that have been obtained for the method of steepest descent [13], [14] are given below.

Theorem 3

If the following assumptions are made

1) $J(u)$ is bounded below
2) $J(u)$ and $g(u)$ are continuous
3) $D^2J(u, h, h)$ exists and

$$|D^2J(u, h, h)| \leq m\|h\|_u^2, \; h \in H_1 \text{ and } m > 0$$

4) $\{u_k\}$ has a cluster point, u^*

then the sequence $\{u_k\}$ formed with arbitrary u_0 by the conjugate gradient method has the following properties:

1) $\lim_{k\to\infty} J(u_k) = J(u^*)$;
2) $\lim_{k\to\infty} g(u_k) = g(u^*) = 0$.

Proof: From the mean value theorem

$$J(u_k + \alpha s_k) = J(u_k) + \alpha(g(u_k), s_k) + \tfrac{1}{2}\alpha^2 D^2J(\xi_k(\alpha), s_k, s_k) \tag{24}$$

where $\xi_k(\alpha)$ belongs to the line segment joining u_k and $u_k + \alpha s_k$. Then Theorem 1 and assumption 3 above give

$$J(u_k + \alpha s_k) \leq J(u_k) - \alpha\|g(u_k)\|^2 + \tfrac{1}{2}\alpha^2 m\|s_k\|^2 \tag{25}$$

$$= J(u_k) - \alpha\|g(u_k)\|^2 \cdot \left(1 - \frac{\alpha m\|s_k\|^2}{2\|g(u_k)\|^2}\right) \forall \alpha, \tag{26}$$

or, assuming that the directions s_k are normalized so that $\|s_k\| = \|g(u_k)\|$

$$J(u_k) - J(u_k + \alpha s_k) \geq \alpha\|g(u_k)\|^2(1 - \tfrac{1}{2}\alpha m). \tag{27}$$

Let

$$\alpha^* = \frac{1}{m}$$

then

$$J(u_k) - J(u_k + \alpha^* s_k) \geq \frac{1}{2m}\|g(u_k)\|^2. \tag{28}$$

Since α_k minimizes $J(u_k + \alpha s_k)$,

$$J(u_k) - J(u_k + \alpha_k s_k) \geq \frac{1}{2m}\|g(u_k)\|^2 \tag{29}$$

and

$$\lim_{k\to\infty} J(u_k) \leq J(u_0) - \sum_{i=1}^{\infty} \frac{1}{2m}\|g(u_i)\|^2. \tag{30}$$

Since $J(u)$ is bounded below,

$$\lim_{k\to\infty} \sum_{i=1}^{k} \frac{1}{2} m\|g(u_k)\|^2 \tag{31}$$

exists and is finite, hence

$$\lim_{k\to\infty} \|g(u_k)\|^2 = 0. \tag{32}$$

By assumption 4, $\{u_k\}$ contains a convergent subsequence $\{\hat{u}_k\}$, with limit point u^*. Then continuity of $g(u)$ implies that $g(u^*) = 0$. From Theorem 2, $J(u_{k+1})$

$< J(u_k)$ and hence by the convergence of u_k and continuity of $J(u)$, property 1 follows.

Computational experience has shown that methods which decrease the function J at each step will generally converge to the nearest local minimum. Since the function is generally convex in some neighborhood of a local minimum, this statement is supported by the following result.

Theorem 4

If the assumptions of Theorem 3 hold and if the following assumption is made:

$$D^2J(u, h, h) \geq M\|h\|^2 \qquad M > 0, \quad u, h \in H$$

then

$$J(u^*) = \underset{u \in H}{\text{Min}}\, J(u).$$

Proof:

$$J(u^* + \alpha p) = J(u^*) + \alpha(g(u^*), p) + \tfrac{1}{2}\alpha^2 D^2 J(\xi(\alpha), p, p) \quad (33)$$

for all $p \in H$ and all α.

By Theorem 3 and the assumption above

$$J(u^* + \alpha p) \geq J(u^*) + \tfrac{1}{2}\alpha^2 M\|p\|^2 \quad (34)$$

or

$$J(u^* + \alpha p) - J(u^*) \geq \tfrac{1}{2}\alpha^2 M\|p\|^2 > 0 \quad \text{if } \alpha \neq 0,\ p \neq 0. \quad (35)$$

IV. Linear Quadratic Case

In order for a minimization technique to perform efficiently on a general function, it is necessary, as has been emphasized previously [5], [6], that it efficiently minimize a quadratic function. This follows since, near the minimum, the quadratic terms in the Taylor series expansion predominate. For the problem of (1), (2), and (3), this corresponds to the linearization of (2) about the minimum and the approximation of (1) by a positive definite quadratic form. The Hilbert space problem becomes the minimization of the quadratic function

$$J(u) = \tfrac{1}{2}(u - u^*, A(u - u^*)) \quad (36)$$

where A is a positive definite self-adjoint linear operator. It is shown by Antosiewicz and Rheinboldt in Todd [6] that there exists a unique $u = u^*$ which minimizes $J(u)$ on H and for which $g(u^*) = 0$.

In addition, it is shown that the conjugate gradient method, when applied to this function, has the following properties:

1) The set $\{s_i\}$ satisfies $(s_i, As_j) = 0\ i \neq j$. This implies that $\{s_i\}$ is a linearly independent set.

2) At the nth iteration, $J(u)$ is minimized over the set

$$B_n = \left\{u \mid u = u_0 + \sum_{j=0}^{n-1} \alpha_j s_j\right\}.$$

Since the s_i are linearly independent, $B_{j-1} \subset B_j$ $j = 1, \cdots, n$ and hence $J(u)$ is minimized over an expanding sequence of sets.

3) The error vectors $y_i = u^* - u_i$ are decreased at each step, i.e., $\|u^* - u_i\| < \|u^* - u_{i-1}\|$ if $g(u_{i-1}) \neq 0$.

4) If the sequence of sets $\{B_n\}$ covers H, that is

$$\bigcup_{n=1}^{\infty} B_n = H$$

then

$$u_i \to u^* \qquad \text{as } i \to \infty.$$

All the above properties, with the exception of 3, hold true for any method which generates conjugate directions when applied to a quadratic function. None of these properties is shared by the method of steepest descent.

In the (finite) n-dimensional case, property 2 implies that the quadratic function will be minimized by any conjugate direction method in at most n steps. This is in contrast to the method of steepest descent, for which it has recently been shown that finite convergence is not obtained unless it is obtained on the first step [15].

A. Further Results

The question of relative performance of the steepest descent and conjugate gradient methods on linear-quadratic problems is of much interest. A number of results follow from the fact that the conjugate gradient method minimizes $J(u)$ over an expanding sequence of sets B_i.

Let barred and unbarred quantities be those achieved by the conjugate gradient and steepest descent methods, respectively. By (15) the sets B_i of property 3 of this section are also given by

$$B_i = \left\{u \mid u = \bar{u}_0 + \sum_{j=0}^{i-1} \alpha_j \bar{g}_j\right\}. \quad (37)$$

Since $u_0 = \bar{u}_0$ and $u_1 = \bar{u}_1$ then $g_1 = \bar{g}_1$ and both u_2 and $\bar{u}_2$ are of the form $u_0 + \alpha_0 \bar{g}_0 + \alpha_1 \bar{g}_1$. Since the conjugate gradient method chooses the α_i to minimize $J(u)$, then

$$J(\bar{u}_2) \leq J(u_2). \quad (38)$$

Similar reasoning applies to the situation where, at step i of the conjugate gradient method, the method of steepest descent is used by setting $\beta_{i-1} = 0$ in (15). The value of J obtained at the next step is never smaller than the value obtained if the conjugate gradient algorithm had been continued.

A more difficult question is whether one method always does better than the other at every iteration, given the same starting point for both. This can be resolved for the problem with one state variable and infinite final time. Consider:

$$\text{minimize} \quad J(u) = \frac{1}{2}\int_0^\infty (x^2 + ku^2)dt \tag{39}$$

$$\text{subject to} \quad \dot{x} = ax + bu, \qquad x(0) = c \tag{40}$$

where all quantities are scalar functions of time.

The relations defining the gradient, (5), (6), and (7), become

$$\dot{x} = ax + bu \qquad x(0) = c, \quad a < 0 \tag{41}$$

$$\dot{\lambda} = -a\lambda - x \qquad \lambda(\infty) = 0 \tag{42}$$

and

$$g(u) = \frac{\partial H}{\partial u} = b\lambda + ku. \tag{43}$$

Assume $u_0 = 0$. [If not, a shift of origin reproduces (41), (42), and (43).] Equations (41) and (42) then both have solutions of the form $c_0 e^{at}$, since the conditions $\lambda(\infty) = 0$ and $a < 0$ allow no term in e^{-at} in the solution of (42). Thus both $g(u_0)$ and u_1 are of this form. Using this u_1 in (41) yields a solution of the form $e^{at}(r_0 + r_1 t)$ for (41) and (42) and $g(u_1)$, u_2, and $\bar{u}_2$ have this form also. A simple inductive argument shows that g_n, $\bar{g}_n$, u_n, and $\bar{u}_n$ are all of the form

$$e^{at}P_{n-1}(t)$$

where P_{n-1} is a polynomial of degree $n-1$ in t. Thus both u_n and $\bar{u}_n$ are elements of the set B_n in (37) and, since $\bar{u}_n$ minimizes $J(u)$ over B_n

$$J(\bar{u}_n) \le J(u_n), \qquad n = 0, 1, \cdots. \tag{44}$$

B. Finite Convergence

It has recently been shown that the steepest descent method applied to a positive semidefinite quadratic function in Hilbert space reaches the optimum in a finite number of steps if and only if it does so on the first step [15]. However, this is not true for the conjugate gradient method, for as the following example shows, finite convergence can be attained in more than one step.

Example: Minimize

$$J(y) = \int_0^1 (\tfrac{1}{2}\dot{y}^2 + y\dot{y} + \dot{y} + y)dt$$

with $y(0) = \frac{1}{2}$. The optimal solution is a quadratic polynomial in t. This can be reformulated as

$$\min_u y_2(1)$$

with

$$\dot{y}_1 = u \qquad y_1(0) = \tfrac{1}{2}$$

$$\dot{y}_2 = \tfrac{1}{2}u^2 + uy_1 + u + y_1 \qquad y_2(0) = 0$$

and optimal solution $u = t - \frac{3}{2}$.

Applying the conjugate gradient method and assuming $u_0 = 0$ leads to

$$s_0 = -\frac{5}{2} + t$$

$$\alpha_0 = 49/97$$

and then

$$u_1 = \frac{99}{194}t - \frac{245}{194}$$

$$g_1 = \frac{-96}{194}t + \frac{44}{194}$$

$$s_1 = \frac{18816}{194 \times 194}t - \frac{9016}{194 \times 194}.$$

The directions s_0 and s_1 are linearly independent and span the space containing the optimal u. Since the conjugate gradient method minimizes $y_2(1)$ over this space, $u_2 = \alpha_0 s_0 + \alpha_1 s_1$ is optimal as is easily verified.

V. Computational Results

A. Introduction

In order to compare the performance of this method with that of steepest descent, three numerical examples are given. The steepest descent technique used is obtained from the conjugate gradient algorithm by setting $\beta_i = 0$ in (15). The linear search was performed using a cubic interpolation technique [6] and is the same for both methods. Thus both methods require about the same computational effort per iteration. Comparison with a second variational method is also given in Example 3.

B. Examples

Example 1. This first example considers the linear system

$$\dot{x}_1 = x_2 \tag{45}$$

$$\dot{x}_2 = -x_2 + u \tag{46}$$

with initial conditions

$$x_1(0) = 0 \tag{47}$$

$$x_2(0) = -1 \tag{48}$$

and quadratic functional

$$J(u) = \int_0^1 \{x_1^2 + x_2^2 + 0.005u^2\}dt. \tag{49}$$

This is the problem considered by Hsieh [16].

Figure 1 shows the variation of the function (49) with number of iterations for the steepest descent and the conjugate gradient methods. The conjugate gradient

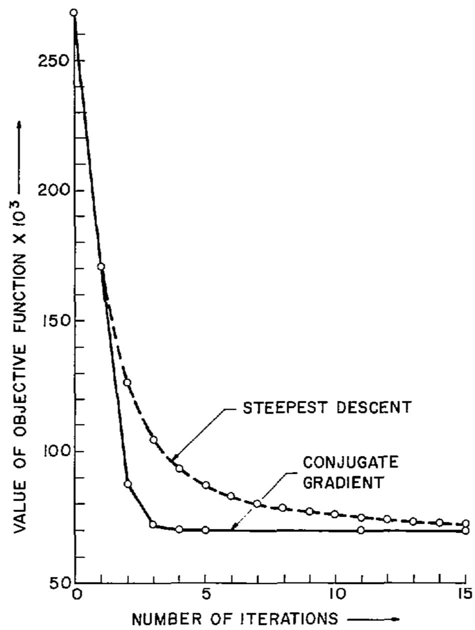

Fig. 1. Behavior of objective—example one.

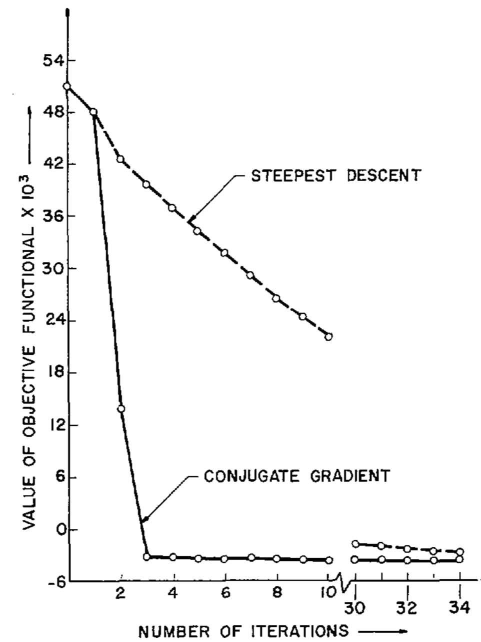

Fig. 3. Behavior of objective—example two.

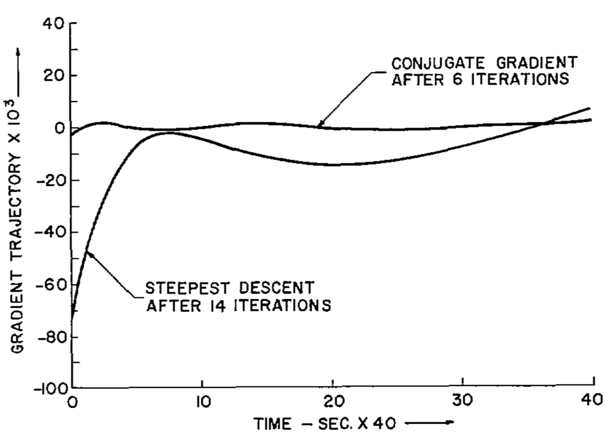

Fig. 2. Gradient trajectories—example one.

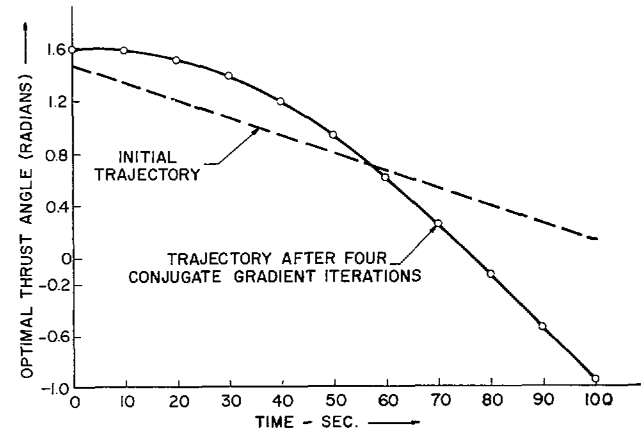

Fig. 4. Optimal control—example two.

technique has effectively converged to the optimum in four iterations, whereas the steepest descent scheme shows slow convergence near the optimum, requiring fifteen iterations to achieve similar results. Figure 2 illustrates the convergence of the gradient trajectories. Obviously, the conjugate gradient method gives superior results.

Example 2. This example is the rocket launch problem described by Dreyfus [17].

$$\dot{x}_1 = x_2 \tag{50}$$

$$\dot{x}_2 = 64 \sin u - 32 \tag{51}$$

$$\dot{x}_3 = 64 \cos u \tag{52}$$

$$x_i(0) = 0 \qquad i = 1, 2, 3 \tag{53}$$

with the objective function

$$\max x_3(100) \tag{54}$$

subject to

$$x_1(100) = 100\,000 \tag{55}$$

$$x_2(100) = 0. \tag{56}$$

Using a penalty function, this problem was converted to a problem with no end-point constraints;

$$\min - x_3(100) + 0.002(x_1(100) - 10^5)^2 + 0.05(x_2(100))^2 \tag{57}$$

subject to (50), (51), (52), and (53).

Figure 3 shows the convergence of the objective function (57) and Fig. 4 the control function obtained by the conjugate gradient method. Gradient trajectories produced are not compared since neither sequence converged, showing instead an oscillatory behavior. This seems to be due to the penalty function terms in (57) which create a steep-sided valley in the control space.

Example 3. As a final example, and one which is nonlinear in the state variables, consider the problem suggested by Merriam [18].

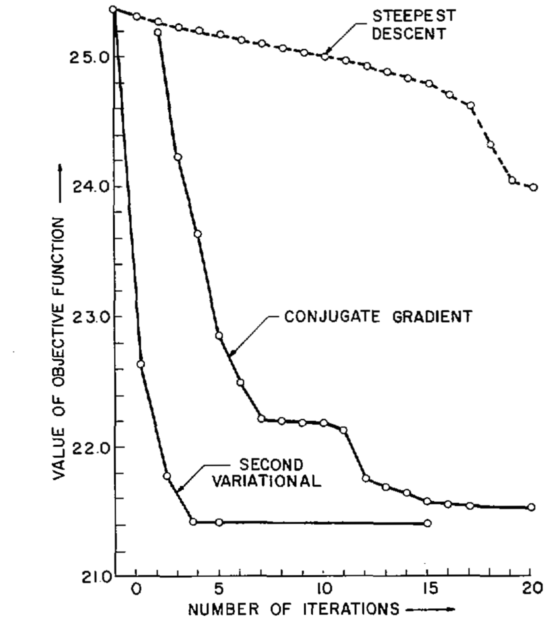

Fig. 5. Behavior of objective—example three.

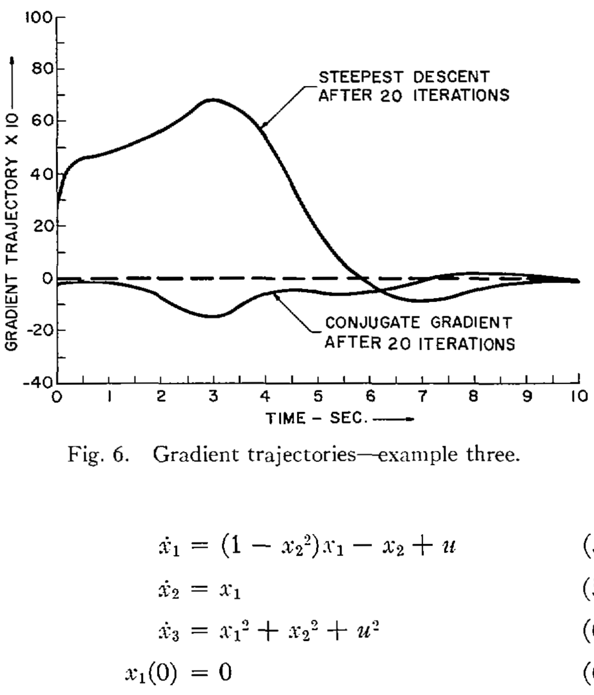

Fig. 6. Gradient trajectories—example three.

$$\dot{x}_1 = (1 - x_2^2)x_1 - x_2 + u \tag{58}$$

$$\dot{x}_2 = x_1 \tag{59}$$

$$\dot{x}_3 = x_1^2 + x_2^2 + u^2 \tag{60}$$

$$x_1(0) = 0 \tag{61}$$

$$x_2(0) = 3 \tag{62}$$

$$x_3(0) = 0 \tag{63}$$

with objective

$$\min x_3(10). \tag{64}$$

Figure 5 compares the convergence of the steepest descent, conjugate gradients, and the second variational method used by Merriam. As in the previous examples the conjugate gradient method is markedly superior to steepest descents. The second variational technique is faster than both but requires considerably more computation per iteration. Figure 6 compares the gradient trajectories for steepest descent and conjugate gradients.

VI. Conclusion

The conjugate gradient method presented here appears to have significant advantages over existing methods. Its convergence rate is evidently superior to that of steepest descents with very little additional computation per iteration. Since the method always generates directions of descent, its stability properties are superior to those of second variational techniques, which may not converge from a poor approximation to the minimum. Second variational methods will probably have superior local convergence, but at greater computational cost. Thus the conjugate gradient method seems to be an effective compromise between existing techniques.

It is apparent that many of the other existing finite dimensional minimization techniques can be similarly extended. For problems with inequality constraints, an extension of the SUMT procedure of Fiacco and McCormick [19] appears promising. In particular, combining SUMT with conjugate gradients should yield a computationally useful algorithm. Such an extension is currently being investigated.

References

[1] R. E. Kopp and R. McGill, "Several trajectory optimization techniques," in *Computing Methods in Optimization Problems*, A. V. Balakrishnan and L. W. Neustadt, Eds. New York: Academic, 1964, pp. 65–89.

[2] H. J. Kelley, "Methods of gradients," in *Optimization Techniques*, G. Leitmann, Ed. New York: Academic, 1962, ch. 6.

[3] W. F. Denham and A. E. Bryson, "Optimal programming problems with inequality constraints-II: Solution by steepest ascent," *AIAA J.*, vol. 2, pp. 25–34, January 1964.

[4] R. McGill, "Optimal control, inequality state constraints, and the generalized Newton-Raphson algorithm," *J. SIAM on Control*, ser. A, vol. 3, no. 2, pp. 291–298, 1965.

[5] R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization," *British Computer J.*, pp. 163–168, June 1963.

[6] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *British Computer J.*, pp. 149–154, July 1964.

[7] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Research NBS*, vol. 49, p. 409, 1952.

[8] H. J. Kelley, "Gradient theory of optimal flight paths," *Am. Rocket Soc. J.*, vol. 30, pp. 947–953, October 1960.

[9] S. R. McReynolds and A. E. Bryson, Jr., "A successive sweep method for solving optimal programming problems," *Proc. JACC*, 1965.

[10] S. Mitter, "Successive approximation methods for the solution of optimal control problems," *Automatica*, vol. 3, pp. 133–149, 1966.

[11] M. J. Box, "A comparison of several current optimization methods and the use of transformations in constrained problems," *The Computer J.*, vol. 9, pp. 67–78, May 1966.

[12] J. Todd, Ed., *Survey of Numerical Analysis*. New York: McGraw-Hill, 1962.

[13] J. R. Rice, *The Approximation of Functions*. Reading, Mass.: Addison-Wesley, 1964, vol. 1.

[14] H. B. Curry, "The method of steepest descent for nonlinear minimization problems," *Quart. Appl. Math.*, vol. 2, pp. 258–261, October 1944.

[15] W. E. Langlois, "Conditions for termination of the method of steepest descent after a finite number of iterations," *IBM J. Research and Dev.*, pp. 98–99, January 1966.

[16] C. T. Leondes, Ed., *Advances in Control Systems*. New York: Academic, 1964–1966, vol. 2.

[17] S. E. Dreyfus, "Variational problems with state variable inequality constraints," RAND Rept. P-2605-1, August 1963.

[18] C. W. Merriam, III, "Direct computational methods for feedback control optimization," *Information and Control*, vol. 8, pp. 215–232, April 1965.

[19] A. V. Fiacco and G. P. McCormick, "The sequential unconstrained minimization technique for nonlinear programming, a primal-dual method," *Management Science*, vol. 10, pp. 360–366, January 1964.