NGR-05-018-022

COMPUTING METHODS IN OPTIMIZATION PROBLEMS

GRADIENT METHODS FOR THE OPTIMIZATION OF DYNAMIC* SYSTEM PARAMETERS BY HYBRID COMPUTATION

George A. Bekey and Robert B. McGhee

Electrical Engineering Department
University of Southern California
Los Angeles, California

I. INTRODUCTION

This paper is concerned with the computer implementation of both continuous and discrete gradient methods for adjusting the parameters of a dynamic system so as to match a specified response function as closely as possible. While the basic theory of parameter optimization by gradient descent has been known for some time, the limitations and convergence properties of particular methods of computer implementation are not yet well understood. This paper is intended to be a contribution toward obtaining a better understanding of these problems.

Continuous parameter optimization is an appealing concept and a number of "adaptive control" schemes have been based on it. The first part of this paper reviews the formulation of a continuous steepest descent algorithm and discusses its difficulties. Computer results relating to the nature of the gradient and the dependence of the path in parameter space on adjustment gain are given.

The second part of the paper reviews briefly several discrete gradient optimization techniques. An algorithm for automatic adjustment of step size for gradient descent is presented. The stability and convergence properties of first and second order iteration schemes are compared and some new results are presented in the form of a convergence theorem. The application of discrete parameter optimization methods to a nonlinear dynamic system is illustrated with an example.

* This research was supported in part at the University of Southern California under U. S. Air Force Office of Scientific Research Grant No. AF-AFOSR-75-63 and in part at Space Technology Laboratories, Inc., under National Aeronautics and Space Administration Contract NAS1-2582.

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
Springfield, Va. 22151

N66 24113 (ACCESSION NUMBER) 23 (PAGES) CR-74896 (NASA CR OR TMX OR AD NUMBER) (THRU) 1 (CODE) 08 (CATEGORY) FACILITY FORM 602

23 pp.

N O T I C E

THIS DOCUMENT HAS BEEN REPRODUCED FROM THE BEST COPY FURNISHED US BY THE SPONSORING AGENCY. ALTHOUGH IT IS RECOGNIZED THAT CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED IN THE INTEREST OF MAKING AVAILABLE AS MUCH INFORMATION AS POSSIBLE.

The final section of the paper presents a formulation of a hybrid computational strategy for parameter optimization which includes the best features of both the analog and digital computer solutions.

2. CONTINUOUS PARAMETER OPTIMIZATION

We consider continuous dynamic systems described by

$$\dot{\bar{y}} = F(\bar{y},\ t;\ \bar{p}) \tag{1}$$

where $\bar{y}$ is an n-vector representing the state of the system and $\bar{p}$ is an m-vector representing the parameters to be optimized, including initial conditions. The parameter optimization problem under consideration is that of selecting $\bar{p}$ in such a way that the solution of Eq. (1) approximates a given function, $y_d(t)$, as closely as possible. The particular criterion function to be used as a basis for parameter adjustment in this paper is given by

$$\phi(\bar{p}) = \int_0^T [y(t;\bar{p}) - y_d(t)]^2\, dt \tag{2}$$

Gradient methods of optimization are based upon adjustment of parameters utilizing the local gradient vector. That is, a parameter change vector, $\Delta\bar{p}$, is computed according to the rule

$$\Delta\bar{p} = -K\,\nabla\phi(\bar{p}) \tag{3}$$

where K is a positive definite matrix and $\nabla\phi$ is the column vector

$$\nabla\phi(\bar{p}) = \left[\frac{\partial\phi}{\partial p_1},\ \frac{\partial\phi}{\partial p_2},\ \cdots \frac{\partial\phi}{\partial p_m}\right]' \tag{4}$$

Following the i-th such calculation, the value of the parameter vector is given by

$$\bar{p}^{(i+1)} = \bar{p}^{(i)} + \Delta\bar{p}^{(i)} \tag{5}$$

The convergence properties of several iteration schemes of this type are discussed in Section 3 of this paper.

Consider now the case where continuous parameter adjustment is desired. It is clear that the criterion function defined by Eq. (2) cannot be used directly since it leads to an iterative adjustment algorithm. Let us therefore define an instantaneous performance criterion

$$f = \frac{d\phi}{dt} = [y(t;\bar{p}(t)) - y_d(t)]^2 \qquad (6)$$

Unfortunately, f is a functional in $\bar{p}(t)$ rather than an ordinary function. Consequently, the gradient vector, $\nabla f(\bar{p})$, does not exist unless $\bar{p}(t)$ is a constant. But this contradicts the original objective of the formulation, namely, to adjust $\bar{p}$ continuously as a function of time. Two different approaches to the resolution of this dilemma have been taken. If the desired output is a vector of derivatives, $\bar{y}_d(t)$, with dimension equal to the order of the system to be optimized, then Eq. (1) may be used to derive a criterion function which is a simple function of $\bar{p}$ even when $\bar{p}$ varies with time. Specifically, if

$$f_e(\bar{p}) = F^2(\bar{y}_d, t; \bar{p}) \qquad (7)$$

then the gradient

$$\nabla f_e(\bar{p}) = 2F \frac{\partial F}{\partial \bar{p}} \qquad (8)$$

exists and may be used to find a minimizing value for $\bar{p}$ by making use of the adjustment algorithm

$$\dot{\bar{p}} = -k \nabla f_e(\bar{p}) \qquad (9)$$

This method, sometimes called the "equation error method" has been used by Graupe (1), Ornstein (2), and others in connection with identification problems.

While the equation error method avoids the difficulty associated with Eq. (6), computer implementation of the method requires that desired values for all of the system state variables be available. An alternate formulation, based on the work of Whitaker (3) and Margolis (4) does not require complete specification of the desired state, but leads only to an approximate gradient method. The degree of approximation is related to the rate of change of adjustment of the parameters as compared to the natural frequencies of both the system and the input process. The remainder of this paper is restricted to the latter formulation; i.e., to circumstances where y_d is a scalar function. While the basic technique to be described for continuous parameter adjustment is not new, the results pertaining to the dynamic properties of the parameter adjustment process have not been previously published.

2.1 The Approximate Gradient Method.

The performance criterion Eq. (7) requires complete knowledge of the desired state. Let us consider instead

the criterion

$$f_c = (e_{c1} + q_1 e_{c2} + \cdots + q_{p-1} e_{cp})^2 \qquad (10)$$

where $p < n$, n is the order of the system, and

$$e_{ci} = y_i - y_{di} = \frac{d^{(i-1)}y}{dt^{(i-1)}} - \frac{d^{(i-1)}y_d}{dt^{(i-1)}} \qquad (11)$$

When $y_d(t)$ is given as a scalar function, error derivatives must be obtained by analog computer differentiation. In many practical situations it is possible to choose all the $q_i = 0$, so that only the system output (or zero-state) is required. The quantity e_c represents "output error" and parameter optimization based on Eq. (10) may be called the "output error method".

Let us choose

$$f_c = (e_{c1} + q\, e_{c2})^2 \qquad (12)$$

Then, if the parameters are constant, the components of the gradient are given by

$$\frac{\partial f_c}{\partial p_i} = 2(e_{c1} + q\, e_{c2}) \frac{\partial}{\partial p_i}(e_{c1} + q\, e_{c2}) \qquad (13)$$

$$i = 1,2, \cdots m$$

Using the definition of e_{c1} and e_{c2} from Eq. (11) and since y_d is independent of the parameters, Eq. (13) can be written as

$$\frac{\partial f_c}{\partial p_i} = 2(e_{c1} + q\, e_{c2}) \frac{\partial}{\partial p_i}(y_1 + q\, y_2); \qquad (14)$$

$$i = 1,2, \cdots m$$

where y_1 and y_2 represent the system output and its first derivative respectively. Let us denote the <u>influence coefficients</u> by the letter u so that

$$u_{ij} = \frac{\partial y_i}{\partial p_j} \qquad (15)$$

The influence coefficients can be obtained by differentiation of the system Eq. (1) with respect to the appropiate parameters and solving the resulting differential equation in u_{ij} (the "sensitivity equation") (5). Analog computer circuits can be used for the simultaneous evaluation of the u_{ij} and the y_i.

Now, if the parameters are adjusted, y_i becomes a

functional and the $\partial y_i/\partial p_j$ do not exist in the ordinary sense. Let us assume, however, that in a given solution interval, the variation is sufficiently slow so that p_j can be assumed constant. Then, a continuous gradient method is based on

$$\dot{p}_i = - k \frac{\partial f_c(\bar{p})}{\partial p_i} \tag{16}$$

However, analog computer circuits based on Eq. (16) are in fact mechanizations of

$$\dot{p}_i = - k\, g_i(\bar{e}_c, \bar{y}, \bar{p}) \tag{17}$$

where the vector $\bar{G} = [g_1, g_2, \cdots g_n]'$ is an approximation to ∇f_c which approaches ∇f_c as $k \to 0$. The u_{ij} which enter into the calculation of the g_i can be considered subsidiary variables which equal the desired sensitivity coefficients when $\bar{p} = 0$.

An analog computer implementation of the approximate gradient method (the output error method) is shown in Fig. 1 for $q = 0$ in Eq. (12). This figure illustrates the application of the method to an identification problem. As long as the switch S is open, the parameters are constant and $\bar{G} = \nabla f_c$. Consequently, the nature of the gradient can be studied in the open-loop case. Then, the switch can be closed for examination of the actual parameter adjustment path.

2.2 The Nature of the Criterion Surface

Let

$$f_c = e_{c1}^2 = (y_1 - y_{d1})^2 \tag{18}$$

and define the parameter offsets δp_i by

$$\delta p_i = p_i^{(o)} - p_i^{(f)}, \quad i = 1, 2, \cdots m \tag{19}$$

where $p_i^{(o)}$ represents the assumed initial values of the parameters and $p_i^{(f)}$ the values which minimize f_c. Then, if the loop is open, we can expand f_c as follows:

$$f_c(t) = \left[e_c(\bar{p}^{(o)}) + \sum_{i=1}^{m} \frac{\partial e_c}{\partial p_i} \delta p_i + O(\delta p^2) \right]^2 \tag{20}$$

If the δp_i are sufficiently small, second and higher order terms may be neglected and

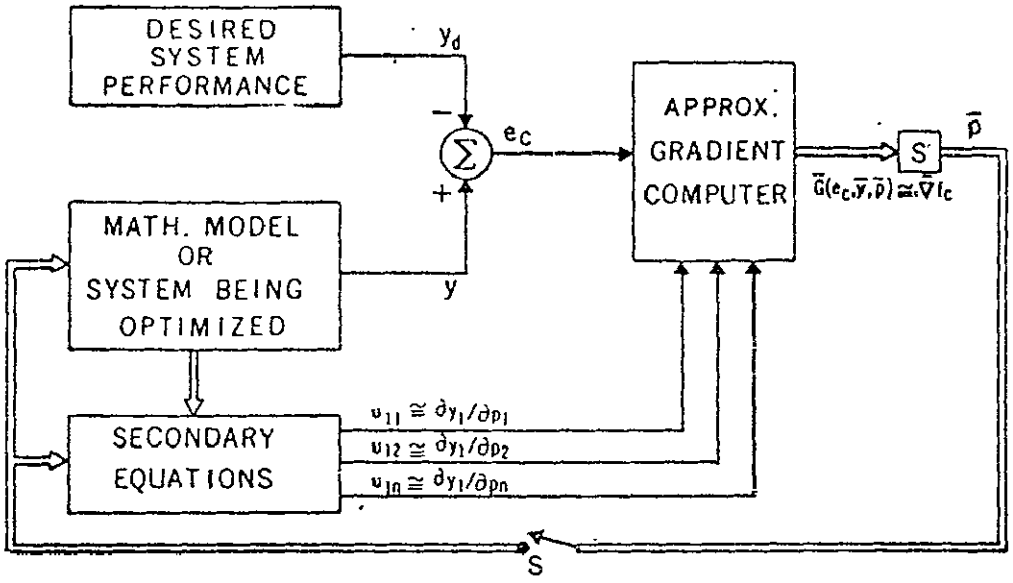

Fig. 1. Schematic of Continuous Parameter Optimization Method

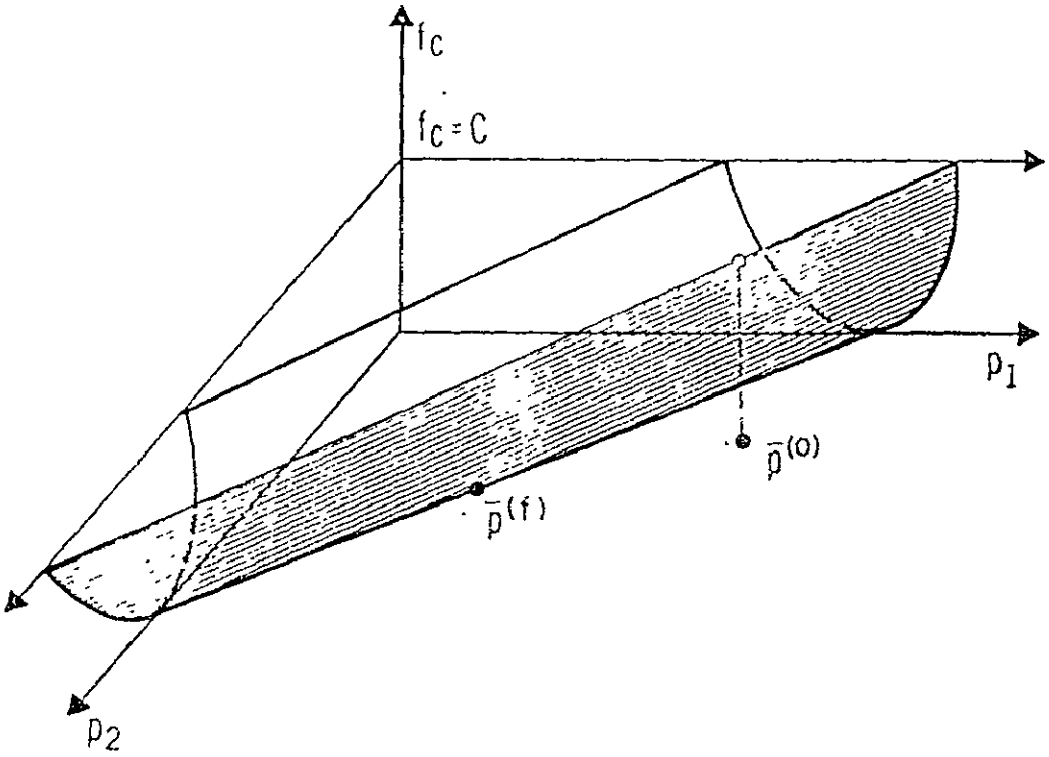

Fig. 2. Instantaneous Criterion Surface

$$f_c(t) = \left[e_c(\bar{p}^{(o)}) + \sum_{i=1}^{m} u_{1i}\ \delta p_i \right]^2 \cdot \qquad (21)$$

Consequently, contours of constant $f_c(t) = C$, at any time t_j, can be found from

$$\sum_{i=1}^{m} u_{1i}(t_j)\ \delta p_i = \pm\ C^{\frac{1}{2}} - e_c(\bar{p}^{(o)}, t_j) \qquad (22)$$

This equation represents two parallel lines in the $m+1$ dimensional space of the parameters and criterion function. If only 2 parameters p_1 and p_2 are present, it is possible to obtain a simple geometrical interpretation of this equation as shown in Fig. 2. It can be seen that the instantaneous criterion function surface is a parabolic trough, where the initial and final parameter values are indicated. The contour lines in the $p_1 - p_2$ plane are straight lines, while the intersection of the trough with the $f_c - p_i$ planes results in the familiar quadratic shape. It should be noted that Fig. 2 represents an instantaneous situation. As the u_{1j} and e_c change with time, the trough moves in such a way that its minimum still crosses the desired final point (6).

2.3 The Gradient Vector

Considerable insight into the nature of the adjustment process is gained if the gradient (with S open, of course) is evaluated with a sinusoidal input. The gradient is given by

$$\bar{\nabla} f_c = 2\left[e_c\, u_{11},\ e_c\, u_{12},\ \cdots\ e_c\, u_{1m} \right]' \qquad (23)$$

Consider, for example, a desired response function obtained from a second order system described by the relation

$$\dot{\bar{y}}_d = A_d \bar{y}_d + B_d \bar{x} \qquad \bar{y}_d(0) = \bar{y}_{do} \qquad (24)$$

where

$$A_d = \begin{bmatrix} 0 & 1 \\ -a_2 & -a_1 \end{bmatrix}, \quad B_d = \begin{bmatrix} 0 & 0 \\ a_3 & a_4 \end{bmatrix}, \quad \bar{x} = \begin{bmatrix} x \\ \dot{x} \end{bmatrix}$$

and the coefficients a_1, a_2, a_3, a_4 are constants. The signal $x(t)$ is the input to the process. It is desired to optimize the parameters α_1 to α_4 of a model described by

$$\dot{\bar{y}} = A\bar{y} + Bx \quad , \qquad \bar{y}(0) = \bar{y}_o \qquad (25)$$

where

$$A = \begin{bmatrix} 0 & 1 \\ -\alpha_2 & -\alpha_1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 \\ \alpha_3 & \alpha_4 \end{bmatrix}$$

using the criterion function of Eq. (18). The locus of the gradient vector can be plotted using the computer in the parameter plane defined by any two of the parameters. To further simplify the visualization of the results, let $\bar{y}_{do} = \bar{y}_o$ and $A_d = A$ so that the differences between y_d and y are due entirely to differences between B_d and B. To compute the gradient (as defined by Eq. (23) the sensitivity coefficients $u_3 = \partial y/\partial\alpha_3$ and $u_4 = \partial y/\partial\alpha_4$ will be required. These coefficients are obtained from computer solution of two subsidiary equations, derived from differentiation of Eq. (25) with respect to α_3 and α_4 respectively. The sensitivity equations for this case are

$$\begin{aligned} \dot{\bar{u}}_3 &= A\bar{u}_3 + C_3\bar{x} \\ \dot{\bar{u}}_4 &= A\bar{u}_3 + C_4\bar{x} \end{aligned} \qquad (26)$$

where

$$C_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \qquad C_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\bar{u}_3 = \begin{bmatrix} u_3 \\ \dot{u}_3 \end{bmatrix} \qquad \bar{u}_4 = \begin{bmatrix} u_4 \\ \dot{u}_4 \end{bmatrix}$$

Substitution of the solution of Eq. (26) into Eq. (23) yields the instantaneous values of the gradient vector. Typical results are shown in Fig. 3, where $x(t)$ is a sinusoid with a frequency of 1 rad/sec. Since the adjustment loop is open, the parameters remain constant, but the sensitivity coefficients and the matching error e_c vary with time, resulting in the Lissajous-like contours in the figure. Since for sinusoidal inputs and linear systems both y_d and y are sinusoidal, the error e_c is also sinusoidal and becomes zero every half-cycle. From an examination of this figure, it is clear that if it is attempted to adjust parameters with a velocity proportional to the gradient, the motion will be oscillatory and may instantaneously point in an erroneous direction.

2.4 Paths of Parameter Adjustment

The dependence of the parameter adjustment path on

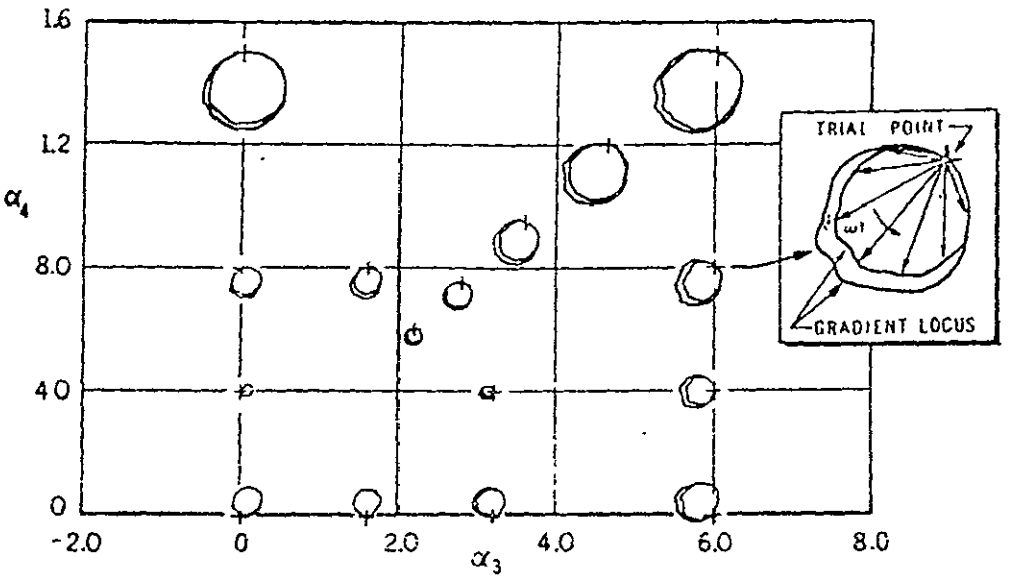

Fig. 3. Open Loop Gradient Loci in the α_3, α_4 Plane

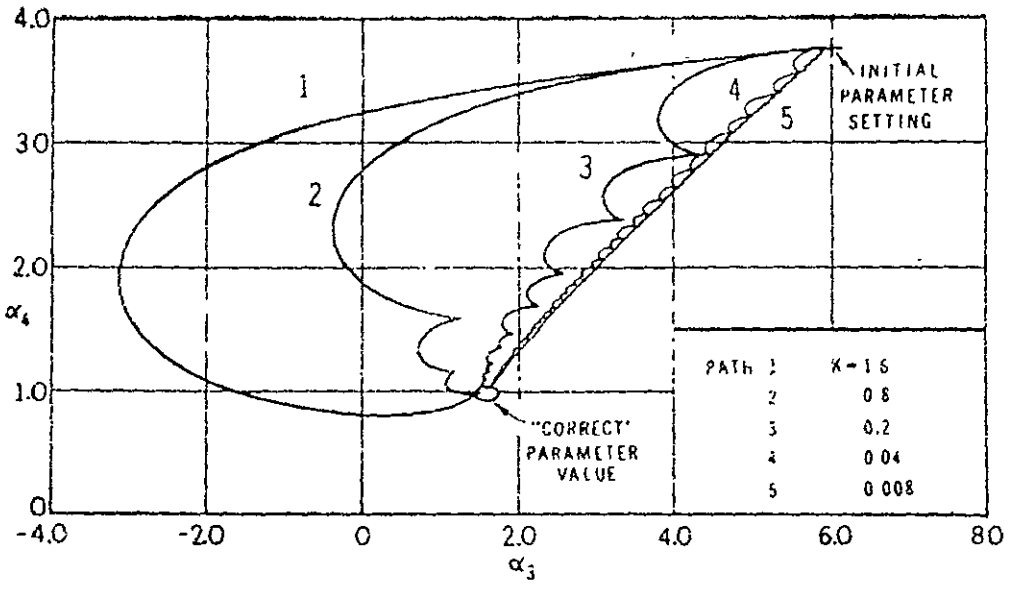

Fig. 4. Descent Trajectories in α_3, α_4 Plane-Sinusoidal Excitation

the parameter k in Eq. (17) is indicated in Fig. 4 for the example discussed previously. When k is very small, the path closely approximates a gradient trajectory. When k is large, the functions g_i are not equal to the respective coefficients of the gradient, but approach a gradient path as the $\delta p_i \to 0$. The "scallops" on the trajectories are due to the oscillatory nature of the approximate gradient vector. Values of k larger than those indicated in this figure may cause instability in the parameter optimization loops.

2.5 Stability of the Parameter Adjustment Loops

General analytical results on stability are not available at the present time. Stability in the small has been demonstrated by Margolis (4) for first and second-order systems with step inputs. Experiments using analog computers show that it is generally possible to find a value of k for which stability and convergence of two or three parameters is possible. However, attempts to improve convergence by increasing k or attempts to adjust more than three parameters simultaneously generally result either in instability or in lack of convergence (6).

At the present time a general existence theorem insuring local stability of the parameter optimization technique for sufficiently small gain is lacking. Such a theorem would prove that a value of k can be found in any particular case such that, for specified classes of inputs and initial conditions, both stability and convergence can be assured.

3. DISCRETE PARAMETER OPTIMIZATION

3.1 Discrete Gradient Descent

The convergence problems encountered in continuous parameter variation schemes may be largely circumvented by making use of a discrete iterative adjustment algorithm. When this is done it becomes possible to determine the true gradient of a given criterion function since parameter changes are made only at discrete points in time. For example, the criterion function given by Eq. (2) may be differentiated to produce

$$\nabla\phi(\bar{p}) = \int_0^T 2\left[y(t;\bar{p}) - y_d(t)\right] \nabla y(t;\bar{p})\, dt \qquad (27)$$

Since $\bar{p}$ is a constant over the interval of integration,

the partial derivatives of y appearing in this expression may be obtained by solving the parameter influence differential equations associated with the assumed equation for y.

The gradient vector computed from Eq. (27) may be used for discrete gradient descent employing Eq. (3). As in the continuous case, the convergence of this parameter adjustment procedure depends upon the values chosen for the elements of K. However, since the gradient used here is exact rather than approximate, discrete gradient descent may always be stabilized by multiplying every element of K by a sufficiently small scale factor. At the present time, an analogous statement cannot be made for continuous parameter adjustment procedures utilizing approximate gradients.

3.2 The Optimum Gradient Method

Since the criterion function, ϕ, is bounded from below by the value zero, it follows that in any region where ϕ is continuous there must exist at least one value for a scalar scale factor k, say $k = k^*$, such that whenever $|\nabla\phi| \neq 0$,

$$\min_{k > 0} \phi\left[\bar{p}^{(i)} - k\,\nabla\phi^{(i)}\right] = \phi\left[\bar{p}^{(i)} - k_i^*\,\nabla\phi(\bar{p}^{(i)})\right] \qquad (28)$$

The "optimum gradient method" (7) utilizes k^* in Eq. (3) i.e., the matrix K is computed anew at every cycle of iteration as

$$K_i = k_i^* I \qquad (29)$$

This choice for K not only guarantees that the sequence of values for ϕ converges, but also assures that the steps taken in parameter space are large enough to make the convergence reasonably rapid.

With the restriction that only a finite number of values for k may be considered, the search for k^* may be carried out automatically by a digital computer. One method for accomplishing this can be based upon an initial value for k computed from the Newton-Raphson formula. If this gain value is called k^0, then at the i-th stage of iteration

$$k_i^0 = \frac{\phi(\bar{p}^{(i)})}{|\nabla\phi(\bar{p}^{(i)})|^2} \qquad (30)$$

and

$$\bar{\Delta p}^{(i)} = -k_i^o \bar{\nabla}\phi(\bar{p}^{(i)}) = -\frac{\phi(\bar{p}^{(i)})\bar{\nabla}\phi(\bar{p}^{(i)})}{\bar{\nabla}\phi(\bar{p}^{(i)}) \cdot \bar{\nabla}\phi(\bar{p}^{(i)})}. \tag{31}$$

Starting with this gain, a binary scale factor search may be conducted by determining an integer $n \geq 0$ which at least locally minimizes the expression

$$\phi(\bar{p}^{(i)}, n) = \phi(\bar{p}^{(i)} - 2^{-n} k_i^o \bar{\nabla}\phi) \tag{32}$$

3.3 A Computational Algorithm for Scale Factor Adjustment

A program to implement scale factor adjustment using the approximate optimum gradient method described above has been written and tested (8). Fig. 5 is a flow-chart for this program *. This algorithm includes a quadratic interpolation formula to permit more accurate determination of an optimum scale factor.

In order to make efficient use of the binary search part of the algorithm illustrated by Fig. 5, it is important that the search begin at a good value for n. Whereas the full Newton-Raphson step ($n=0$) obtained from Eq. (31) may produce good convergence in the early phases of iterative optimization, it has been found in numerical experiments that ever larger values of n are needed in the terminal stages unless ϕ attains the value zero at its minimum (8). This comes about because Newton-Raphson iteration is based upon linear extrapolation of ϕ to zero. When $\bar{\nabla}\phi$ approaches zero while ϕ remains positive, the computed parameter change vector grows without bound unless there is a corresponding increase in the value of n used in the binary scale factor of Eq. (32). Computational experience indicates that this difficulty may be resolved by beginning each scale factor search with the value for n which was found to be optimum during the previous search cycle (8).

3.4 Constrained Minimization

While application of the optimum gradient method to parameter optimization problems does indeed lead to a convergent sequence in ϕ, it cannot be assumed that the sequence in p also converges. This difficulty can be

---

* The numbers on the various blocks of this figure refer to FORTRAN statement numbers. In this diagram, the symbol c rather than $\bar{p}$ has been used to represent a parameter vector.

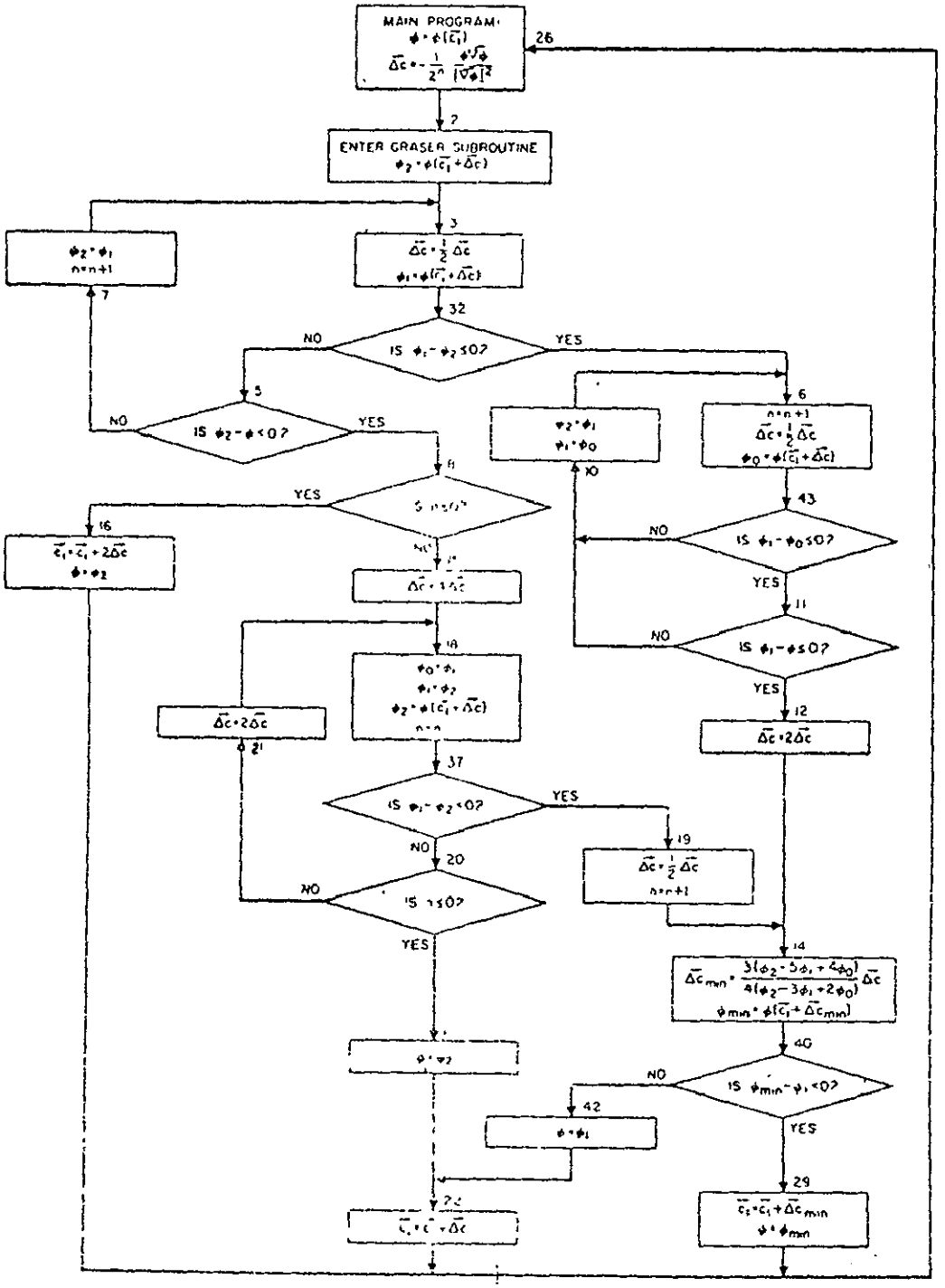

Fig. 5. An Algorithm for Determining the Optimum Scale Factor for Gradient Descent

avoided by specifying a bounded set from which $\bar{p}$ must be chosen. The "gradient projection" method devised by J.B. Rosen can be applied to closed convex constraint sets to obtain constrained minima by gradient searching (9,10). A very much simpler procedure has been developed to permit the use of the algorithm given in this paper in conjunction with independently constrained parameters (8); i.e. with constraints of the form

$$a_k \leq p_k \leq b_k \tag{33}$$

Application of such constraints often produces better convergence even in optimization problems not naturally constrained (8).

3.5 Second Order Methods

The convergence of iterative minimization procedures can be sharpened markedly in certain circumstances by making use of second derivative information. In particular, "Newton's" method, given by

$$\Delta\bar{p}^{(i)} = \left[\frac{\partial^2 \phi}{\partial p_i \partial p_j}\right]^{-1} \nabla\phi(\bar{p}^{(i)}) \tag{34}$$

possesses quadratic convergence properties at a regular minimum of ϕ (11).

The matrix of second partials in Eq. (34) can be obtained by differentiating each row of Eq. (27) with the result

$$\left[\frac{\partial^2 \phi}{\partial p_j \partial p_k}\right] = \int_0^T 2\left(\nabla y \; \nabla y' + e\left[\frac{\partial^2 y}{\partial p_j \partial p_k}\right]\right) dt \tag{35}$$

$$= 2(S - D) \tag{36}$$

where

$$S = \int_0^T \nabla y \; \nabla y' \, dt \tag{37}$$

$$e(t;\bar{p}) = y(t;\bar{p}) - y_d(t) \tag{38}$$

$$D = \int_0^T e\left[\frac{\partial^2 y}{\partial p_j \partial p_k}\right] dt \tag{39}$$

The matrix S amounts to a regression matrix since it links the linear dependence of the response function, $y(t;\bar{p})$, to the integral squared error function, ϕ.

To avoid the computational difficulties associated with the determination of second derivatives, it has sometimes been suggested that Newton iteration be modified to include only the S matrix in Eq. (36) (12). If this is done, the basic iteration equation becomes

$$\Delta\bar{p} = -\frac{1}{2} S^{-1} \nabla\bar{\phi}(\bar{p}) \tag{40}$$

This scheme has been called "Gauss-Newton" iteration (13). Depending upon the magnitude of D relative to S, it may not converge. More precisely, it is shown in the appendix to this paper that if ϕ possesses a regular minimum at $\bar{p} = \bar{p}_o$, then if S_o and D_o denote the values for S and D at $\bar{p}_o$, a region of convergence for Gauss-Newton iteration exists about $\bar{p}_o$ if and only if all of the eigenvalues of the matrix

$$Q = S_o^{-1} D_o \tag{41}$$

are less than one in absolute value.

The residual error, e, existing at a minimizing value for $\bar{p}$, $\bar{p} = \bar{p}_o$, appears as a multiplicative factor in the expression for D. Consequently, the eigenvalues of Q will tend to be small when this error is small and large when it is large. When $y_d(t)$ represents a function which can be matched exactly by a solution of the assumed system equation, Eq. (1), then at $\bar{p} = \bar{p}_o$, e is identically zero and all of the eigenvalues of Q are likewise equal to zero. Gauss-Newton iteration reduces to Newton iteration in this circumstance and quadratic convergence is obtained utilizing only the first derivative information contained in S

3.6 An Example of Discrete Parameter Optimization

The quadratic convergence predicted for Gauss-Newton iteration has been observed in numerical experiments (8). These experiments involved optimization of the four dimensional parameter vector associated with the nonlinear differential equation

$$c_2\ddot{y} + c_1\dot{y} + \sin y = 0$$

$$y(0) = c_3, \quad \dot{y}(0) = c_4 \tag{42}$$

The desired response, $y_d(t)$, was obtained from numerical solution of this equation with a given value for $\bar{c}$. An incorrect value was then taken as a starting point and the computer was permitted to adjust this value itera-

tively using both the optimum gradient and Gauss-Newton methods. Tables 1 and 2 summarize the results of this experiment. Table 2 shows that quadratic convergence is indeed obtainable through Gauss-Newton iteration even though only first derivative information is used.

4. HYBRID COMPUTER IMPLEMENTATION

4.1 Division of the Computation Load Between Analog and Digital Machines

In principle, any of the methods which have been described could be implemented on either a digital computer or an iterative analog computer. When realistic equipment limitations are taken into account however, continuous parameter adjustment is most naturally carried out by analog computation while iterative adjustment seems to be best suited to a digital computer. Both of these choices suffer from certain drawbacks, however. As has been noted, the stability of a continuous parameter adjustment algorithm is very difficult to ascertain a priori. Generally, manual intervention is required to achieve a loop gain producing reasonably rapid convergence without instability. On the other hand, when a completely digital solution to dynamic system parameter optimization problems is attempted, it is quite likely that an excessive amount of computer time will be required since digital machines are not naturally suited to high speed iterative solution of differential equations. For most optimization algorithms, the best features of both types of machines seem to be needed.

A combination of digital decision and branching capabilities and analog solution speed is available in a hybrid computer. With a hybrid computer possessing a sufficiently flexible control structure, analog computer potentiometers and initial conditions can be adjusted automatically under program control so that the analog machine effectively provides high speed subroutines to the digital computer whenever differential equation solution is required. Conversely, by monitoring the results of continuous parameter adjustment via analog to digital converters, the digital machine can assure stability in otherwise uncertain circumstances. Finally, an appropriate mixture of discrete and continuous parameter optimization algorithms can be used on the same problem under overall digital control. One might, for example, utilize discrete methods in large error conditions and continuous methods for "fine tuning" of parameters.

Table 1. Parameter Optimization by the Optimum Gradient Method

Iteration Number	c_1	c_2	c_3	c_4	ϕ
0	.5100	1.010	1.570	.01000	1.853×10^{-3}
1	.5053	1.005	1.567	.00589	4.662×10^{-4}
2	.5027	1.003	1.566	.00391	1.185×10^{-4}
3	.5012	1.002	1.565	.00295	3.109×10^{-5}
True Values	.5000	1.000	1.569	.0000	0

Table 2. Parameter Optimization by Gauss-Newton Iteration

Iteration Number	c_1	c_2	c_3	c_4	ϕ
0	.51000000	1.0100000	1.5700000	.01000000	1.853×10^{-3}
1	.50013945	.99966252	1.5690003	.00007233	4.428×10^{-7}
2	.50000021	1.0000002	1.5690004	.00000068	1.086×10^{-13}
True Values	.50000000	1.0000000	1.5690000	.0000000	0

4.2 An Algorithm for Parameter Optimization by Hybrid Computation

Figure 6 provides an example of an overall algorithm suitable for implementation on a hybrid computer. The REGRES subroutine appearing on this figure determines $\phi(\bar{c})$ and $\nabla\phi(\bar{c})$ as well as the regression matrix, S, by analog solution of the assumed system equation and the associated parameter influence equations. The Gauss-Newton parameter change vector, designated $\bar{\beta}$, is then evaluated by digital inversion of the S matrix followed by matrix multiplication as in Eq. (40). GRASER is the Fortran Symbolic name attached to the scale factor adjustment routine illustrated by Fig. 5. The numerous evaluations of ϕ required in the execution of the GRASER subroutine are also intended to be accomplished by analog means. However, all of the decisions appearing at branch points of both Figs. 5 and 6 are realized by a digital program.

Both the Gauss-Newton and optimum gradient iteration techniques are incorporated into this algorithm. As Fig. 6 shows, the routine favors the Gauss-Newton parameter change vector, $\bar{\beta}$, and switches to the optimum gradient method during a given iteration cycle only when the Gauss-Newton vector fails to satisfy certain criteria. The algorithm provides for independent constraints on the values of each parameter; the region R refers to the n-dimensional box defined by these constraints.

Since the parameter adjustment procedures used here are all iterative, some means for stopping the iteration must be provided. Fig. 6 incorporates five different stopping rules operating in parallel. The d_ϕ and d_c criteria refer to the percentage change in ϕ and c in two successive iterations. When either of these variables falls below a value specified in advance, computation ceases.

4.3 Experimental Results

The algorithm proposed was tested using the system described by Eq. (42). The desired response function, $y_d(t)$, was obtained by numerical solution of this equation with a specified parameter vector. Since a hybrid computer was not available, digital subroutines were used to simulate the necessary analog computations.

Table 3 summarizes the results of this experiment. The column labeled "n" denotes the exponent of the optimum binary scale factor found by gradient searching (Eq. (32)). The entry G-N in this table indicates that

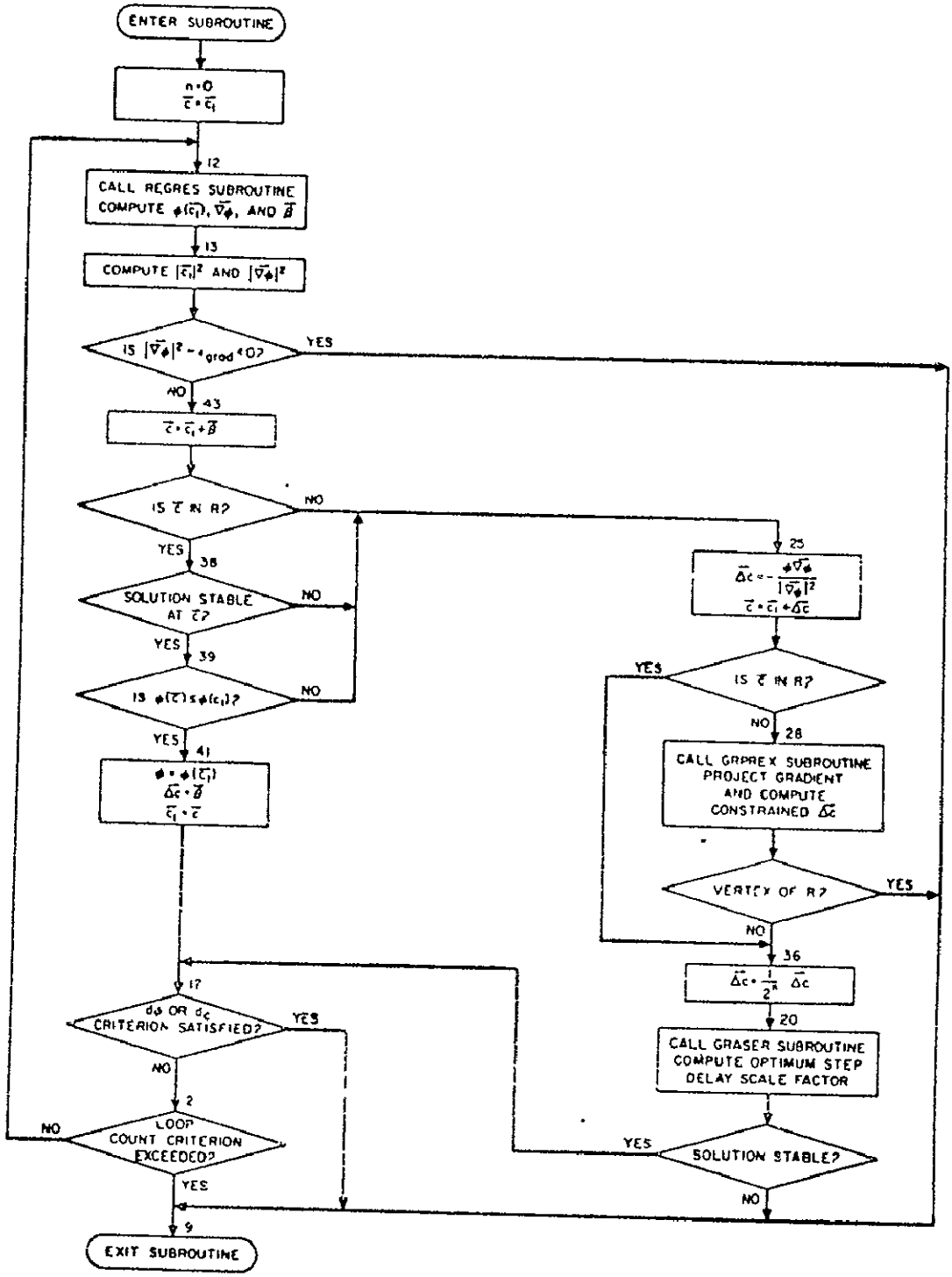

Fig. 6. A Hybrid Computer Parameter Optimization Algorithm

Table 3. Sequence of Parameter Estimation Produced by the Hybrid Computer Algorithm

Iteration Number	c_1	c_2	c_3	c_4	ϕ	n
0	1.0000	2.0000	3.0000	1.0000	360.3	-
1	1.5000	1.8176	2.5322	0.5201	34.12	0
2	1.5000	1.6091	1.6341	-0.2000	1.462	0
3	1.0907	1.4829	1.0000	-0.2000	0.690	0
4	0.8018	1.5123	1.3731	-0.1412	0.366	0
5	0.7581	1.4656	1.3567	-0.2000	0.270	1
6	0.5460	1.2693	1.3978	-0.2000	0.109	0
7	0.5494	0.9792	1.6642	-0.1253	8.48×10^{-3}	G-N
8	0.5000	1.0005	1.5619	-0.0111	2.49×10^{-4}	G-N
9	0.5001	1.0000	1.5691	-0.0001	1.78×10^{-8}	G-N
10	0.5000	1.0000	1.5690	0.0000	3.16×10^{-14}	G-N
True Value	0.5000	1.0000	1.5690	0.0000	0	-
Upper Limit	1.5000	2.5000	3.5000	1.5000	-	-
Lower Limit	0.0000	0.0000	1.0000	-0.2000	-	-

Gauss-Newton iteration rather than gradient searching was used at that step.

5. CONCLUSIONS

Continuous parameter optimization algorithms are conceptually simple but their convergence properties are difficult to determine. Global stability of optimizing algorithms can be obtained by making use of digital computer supervision and control. Since analog computers are naturally suited to the high speed iterative solution of differential equations, hybrid analog-digital computation seems to offer considerable promise as a practical means for system optimization.

APPENDIX

Necessary and Sufficient Conditions for the Existence of a Region of Convergence for Gauss-Newton Iteration

Theorem:

Let $\bar{p}_o$ represent a point in parameter space at which $\phi(\bar{p})$ attains minimum value and let S_o and D_o denote the values of the matrices S and D at this point. Assume that there exists an ϵ-neighborhood about p_o such that $\phi(\bar{p})$, $\nabla\phi(\bar{p})$, $y(t;\bar{p})$, and $\nabla y(t;\bar{p})$ all possess uniformly convergent Taylor series. Suppose further that the matrix S is non-singular everywhere in the same neighborhood. Then, provided that $S_o + D_o$ is a positive definite matrix, when $\Delta\bar{p}(i)$ is computed by the Gauss-Newton formula and $\bar{p}(1)$ is chosen so that

$$|\bar{p}(1) - \bar{p}_o| < \epsilon_o \ , \ 0 < \epsilon_o < \epsilon \tag{43}$$

there exists an ϵ_o such that

$$\lim_{i \to \infty} \bar{p}(i) = \bar{p}_o \tag{44}$$

if and only if all of the eigenvalues of the matrix $Q = S_o^{-1} D_o$ are less than one in absolute value.

Proof:

Let $\delta\bar{p}(i)$ denote the difference between $\bar{p}(i)$ and the desired minimizing value for $\bar{p}$; i.e.,

$$\delta\bar{p}(i) = \bar{p}(i) - \bar{p}_o \tag{45}$$

Then from Eq. (35)

$$\phi(\bar{p}) = \phi(\bar{p}_o) + \delta\bar{p}' (S_o+D_o) \delta\bar{p} + O(\delta p^3) \quad (46)$$

and therefore

$$\nabla\phi(\bar{p}) = 2(S_o + D_o) \delta\bar{p} + O(\delta p^2) \quad (47)$$

Substituting this expression into the Gauss-Newton formula, Eq. (40), produces the result

$$\Delta\bar{p}^{(i)} = - S(\bar{p}^{(i)})^{-1} (S_o+D_o) \delta\bar{p}^{(i)} \quad (48)$$

However, under the assumptions regarding $y(t;\bar{p})$ and $\nabla y(t;\bar{p})$,

$$S(\bar{p}^{(i)}) = S_o + O(\delta\bar{p}^{(i)}) \quad (49)$$

so Eq. (48) reduces to

$$\Delta\bar{p}^{(i)} = (- I - S_o^{-1} D_o + R_i) \delta\bar{p}^{(i)} \quad (50)$$

where R_i is a remainder matrix which tends to zero as $\delta\bar{p}$ approaches zero. Now since

$$\delta\bar{p}^{(i+1)} = \delta\bar{p}^{(i)} + \Delta\bar{p}^{(i)} \quad (51)$$

Eq. (50) yields the recursion relation

$$\delta\bar{p}^{(i+1)} = (- S_o^{-1} D_o + R_i) \delta\bar{p}^{(i)} \quad (52)$$

By choosing ϵ_o sufficiently small, the remainder matrix, R_i may be made as small as desired in comparison to Q. Therefore, for $|\delta\bar{p}^{(1)}| < \epsilon_o$, $\delta\bar{p}$ converges to zero if and only if all of the eigenvalues of Q are less than one in absolute value.

REFERENCES

1. Graupe, K. K., "Analog Solution of Some Functional Analysis Problems", AIEE Transactions, January 1961.

2. Potts, T. F., Ornstein, G. N., and Clymer, A. B., "The Automatic Determination of Human and Other System Parameters", Proceedings of the Western Joint Computer Conference, May 1961, pp. 645-660.

3. Whitaker, H. P., Yamron, J., and Kezer, H., Design of Model Reference Adaptive Control Systems for Aircraft, MIT Instrumentation Laboratory Report R-164, Sept., 1958.

4. Margolis, M., On the Theory of Process Adaptive Control Systems, the Learning Model Approach, Report No. 60-32, University of California, Department of Engineering, Los Angeles, California, (May 1960).

5. Meissinger, H. F., "The Use of Parameter Influence Coefficients in Computer Analysis of Dynamic Systems", Proceedings of the Western Joint Computer Conference, 1960, pp. 181-192 (May 1960).

6. Dekey, G. A., and Meissinger, H. F., and Rose, R. E., "Study of Model Matching Techniques for the Determination of Parameters in Human Pilot Models", Reports No. 8426-6002-RU-000, 8426-6003-RU-000, and 8426-6004-RU-000, Space Technology Laboratories, Redondo Beach, Calif. (1963).

7. Brombery, N. S., "Maximization and Minimization of Complicated Multi-Variable Functions", AIEE Transactions, Vol. 58, Part 1, pp. 725-730, (January 1962).

8. McGhee, R. B., Identification of Nonlinear Dynamic Systems by Regression Analysis Methods, Ph.D. dissertation University of Southern California, Los Angeles, California (June 1963).

9. Rosen, J. B., "The Gradient Projection Method for Nonlinear Programming, Part I, Linear Constraints", J. Soc. Indust. Appl. Math., Vol. 8, No. 1, pp. 181-217 (March 1960).

10. Rosen, J. B., "The Gradient Projection Method for Nonlinear Programming, Part II, Nonlinear Constraints", J. Soc. Indust. Appl. Math, Vol. 9, No. 4, pp. 514-532, (December 1961).

11. Chernoff, H., and Crockett, J. B., "Gradient Methods of Maximization", Pacific Journal of Mathematics, Vol. 5, pp. 33-50 (1955).

12. Williams, E. J., Regression Analysis, John Wiley and Sons, Inc., New York, N. Y., (1959).

13. Hartley, H. O., "The Modified Gauss-Newton Method for the Fitting of Nonlinear Regression Functions by Least Squares", Technometrics, Vol. 3, No. 2, pp. 269-280 (May 1961).

SPACE.

4. Document affected-No. and title: SEE REMARKS

N66-~~24113~~ 24113

2. Date issued: 12 MAY 67

3. Authority for change: ☐ Input ☐ DDC ☐ AID ☐ NASA ☐ Oral ☐ Written; 5. Issued by: ASHLEY SUTTON/CLAY

6. Changes
a. Cancel, duplicate of:
b. Cancel, superseded by:
c. Withdraw from public sale
d. Destroy doc & related papers
e. Return doc & related papers to:
f. Other: ☐
g. Change dist/avail code to:
h. Other:
NOT REPRODUCIBLE, NOT AVAILABLE.

7. Remarks
N66-23818 (NASA CR-71789)
N66-24113 (NASA CR-71806)
N66-24405 (NASA CR-74628)
N66-24404 (NASA CR-74631)
N66-24664 (NASA CR-71694)
N66-25375 (NASA CR-66099)
N66-25563 (NASA CR-65842)

Clearinghouse for Federal Scientific and Technical Information
MAY 67

U.S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

☐ ANNOUNCEMENT ☐ CATALOG
☐ DDC ☐ ARCHIVES
☐ INPUT

1 STOCK LOCATION

2 DATE RECEIVED: YR 72 MO 01 DAY 27

12 SCREEN: ☐ OBTAIN BETTER COPY ☐ OBTAIN AUTHORITY ☐ REJECT ☐ OUT OF PRINT SOD ☐ ERRATA

17 ACCESSION NUMBER: N66-24113

3 RECEIPT TYPE & FORMAT: ☐ MAGNETIC TAPE
☐ LOAN ☒ PC ☐ 35 MM ☐ CARDS
☒ RETAIN ☐ MF ☐ 16 MM ☐ OTHER →

13A ANNOUNCEMENT: VOL 68 ISSUE 26 ☐ SG ☐ TAB ☒ UNANN

13B FAS: ☐ YES ☐ NO

18 PAGES: 27

19 SHEETS: —

20 LOW LIMIT: PC MF

21 ☐ SUB-SCRIP-TION

4 STOCK RECEIVED FOR SALE: PC 0 MF

14 REPRODUCTION INSTRUCTIONS

BLOWBACK →

PRINT ↓	NO	1UP	2UP
NO	1	4	7
1UP	2	5	8
2UP	3	6	9
MIX	SAME SIZE	ORDER STOCK FROM	0

MAKE MICROFICHE: ☐ YES ☒ NO

15 PRESTOCK

	NO!	# COPIES
PC		
MF		0
PC DUE IN		

SOURCE TO ORDER COPIES

22 PRICES: ☒ U UNIT ☐ P PC → ☐ E PC + MN BOX 16 ☐ M MN →

23A CATEGORY

5 LOAN DOCUMENT RETURNED

DUE

OUT

24 DISTR CODE: —

25 INITIALS: ACC EC A B

23B CATEGORY

6 TRANSACTION: NEW ITEM X DUPE ☐ SUPER-SEDES ☐ PRIOR NUMBER ☐

26 FILL FROM: PAPER COPY ETC BX ☐ ↓ MICRO-NEGA-TIVE XM

27 PUBLIC RELEAS-ABILITY: A

7

8 REPORT NUMBERS (XREF) ☐→

16 REMARKS

9 RELATED DOCUMENT ☐→

10 CONTRACTING OFFICE — BILLING CODE: NASA — 000

11 NOT FULLY LEGIBLE ☐ COLOR ☐

28 FORM & PRICE: 1611

FORM NTIS-77 (10/70)

DOCUMENT TRAVELER

NATIONAL TECHNICAL INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE

1 ARCHIVES

3/3