A Comparison of Gradient Dependent Techniques for the Minimization of an Unconstrained Function of Several Variables

TERRY A. STRAETER* AND JOHN E. HOGGE*
NASA Langley Research Center, Hampton, Va.

A simplified aerospace problem and Rosenbrock's parabolic valley function serve as a basis for comparison with respect to ease of implementation and speed of convergence in terms of number of function evaluations. The aerospace problem is a two-dimensional minimum time to intercept problem with constant thrust in free space. The algorithms compared are the 1) Davidon-Fletcher-Powell variable metric, 2) Goldstein's algorithm, 3) Fletcher-Reeves conjugate gradient, and 4) Davidon's second method. Various linear minimization schemes are compared when using the Davidon-Fletcher-Powell variable metric algorithm. The Armstrong-Marquardt method, a minimization algorithm for a weighted sum-of-squares-type function arising in boundary-value problems in optimal control, is also applied to the aerospace example. Tables summarizing the results are given.

Introduction

IN many areas of applied mathematics it becomes necessary to minimize a function of several variables. Examples of these areas are nonlinear programing problems arising in aircraft design and synthesizing optimal trajectories.

In most interesting problems the functions involved are too complicated to be minimized by hand. Some automatic iterative scheme is needed to solve the problem on the digital computer. The algorithm should 1) be simple and reliable, 2) generate a monotonic decreasing sequence of function values, and 3) use as little computer time as possible.

There are many methods of locating the minimum, but basically only two types; that which requires the calculation of the function and its gradient, and that which is not gradient dependent. Gradient dependent techniques will be discussed in this paper. If an analytic expression for the gradient is not readily available, Stewart[1] has shown that in most cases the information available from the Fletcher-Powell[2] technique, based on Davidon's first method,[3] enables one to accurately compute the gradient by differencing.

The techniques investigated are characterized by the fact that they use only first-order information (e.g., the gradient), and yet near the minimum the convergence becomes nearly quadratic. They are the Fletcher-Powell adaptation of Davidon's first method, the conjugate gradient of Fletcher and Reeves,[5] Goldstein-Price,[4] Davidon's second method,[6] and the Armstrong-Marquardt method.[7,8] In this paper a brief description of the methods is given. Then the methods are compared with respect to speed of convergence and ease of computer implementation using a test function and a simple control problem. Recommendations regarding which methods to use on various classes of problems are given. The computations were made on a CDC 6600 computer.

Background

If $f(X)$ has continuous third partials and has a relative minimum at Y, then by Taylor's theorem we have

$$f(X) = f(Y) + \tfrac{1}{2}(X - Y)^T[\partial^2 f/\partial x^2(Y)](X - Y) + \text{h.o.t.} \quad (1)$$

Since Y is the location of a relative minimum of $f(X)$, it is necessary that $\nabla f(Y) = 0$ and $\partial^2 f/\partial x^2(Y)$ is positive semi-definite. If the high-order terms go to zero as $\|X - Y\| \to 0$, and if X is sufficiently close to Y then

$$\nabla f(X) \cong [\partial^2 f/\partial x^2(Y)](X - Y) \quad (2)$$

If the matrix of second partials of $f(X)$ at Y is positive definite, then its inverse exists and is positive definite. Thus from Eq. (2) it follows that

$$Y = X - [\partial^2 f/\partial x^2(Y)]^{-1} \cdot \nabla f(X) \quad (3)$$

Although the matrix of second partials is not known in general, Eq. (3) gives the motivation for most of the minimization algorithms discussed herein. That is, the new estimate of the location of the minimum equals the old estimate minus the product of some positive definite matrix used to approximate $[\partial^2 f/\partial x^2(Y)]^{-1}$ and the gradient of the function at the old estimate.

Suppose $f(X)$ had to be expanded about some other point say X^*. Then

$$f(X) = f(X^*) + (X - X^*)^T\nabla f(X^*) + \tfrac{1}{2}(X - X^*)^T[\partial^2 f/\partial x^2(X^*)](X - X^*) + \text{h.o.t.} \quad (4)$$

Omitting higher order terms and differentiating with respect to X we have

$$\nabla f(X) = \nabla F(X^*) + [\partial^2 f/\partial x^2(X^*)](X - X^*) \quad (5)$$

Table 1 Comparison of gradient methods using the Rosenbrock parabolic valley function $f(x_1,x_2) = (x_2 - x_1^2)^2 + 0.01(x_1 - 1.0)^2$

Method	Iterations	Function evaluations	Gradient evaluations	Value of function of convergence
Fletcher-Powell[a]	25	101	101	0.0
Fletcher-Powell[b]	23	132	132	1.5×10^{-29}
Fletcher-Powell[c]	18	227	18	5.0×10^{-31}
Fletcher-Reeves	30	173	173	3.0×10^{-29}
Goldstein-Price	22	115	115	7.0×10^{-25}
Davidon's second	64	64	64	2.5×10^{-27}

[a] Davidon's cubic interpolation used in the one-dimensional minimization.
[b] Method of false position used in the one-dimensional minimization.
[c] Method of golden sections and cubic fit used in one-dimensional minimization.

Presented as Paper 69-951 at the AIAA Aerospace Computer Systems Conference, Los Angeles, Calif., September 8–10, 1969; submitted Oct. 27, 1969; revision received March 16, 1970.

* Aerospace Technologist.

Now at $X = Y$, the relative minimum $\nabla f(Y) = 0$. Thus from Eq. (5) it follows that

$$Y = X^* - [\partial^2 f/\partial x^2(X^*)]^{-1}\nabla f(X^*) \tag{6}$$

provided of course the Hessian of $f(X)$ is nonsingular at X^*. Equation (6) is the well known Newton-Raphson algorithm.[15] But there are three well recognized problems which may arise using this method. First, the computation of the matrix of second partials may be difficult. Second, away from the minimum, this matrix may not be positive definite or even invertible. And finally, if the starting point is not close to the minimum, the algorithm may not converge or even generate a monotone decreasing sequence of function values.

Algorithms

Fletcher-Powell Adaptation of Davidon's First Method

An algorithm for minimizing a function of several variables was published by W. C. Davidon in 1959.[3] In 1963, R Fletcher and M. J. S. Powell proved theorems about the rate of convergence and the stability of one of the options in Davidon's method. Fletcher and Powell showed that for a quadratic function of n variables the location of the minimum is found after at most n iterations. Moreover, in the algorithm the matrix H^i converges to the inverse of the matrix of second partials after n iterations, and for all iterations H^i remains positive definite. Thus for the general case, the location of the minimum and the inverse of the matrix of second partials are given at convergence. This byproduct is used to good advantage by Stewart and Kelly.[16]

The method replaces the n dimensional minimization problem with a sequence of one-dimensional minimizations. The accuracy of the one-dimensional or linear minimization is crucial to the convergence of the algorithm. The efficiency of this minimization determines the rate of convergence of this method. The algorithm is as follows. Let X^0 be given and $H^0 = I_n$ (or any positive definite matrix): step 1) $S^i = -H^i\nabla f(X^i)$; step 2) the minimum of $f(X)$ along the line $X^i + \alpha S^i$ is found by some technique; step 3) the point found in step 2 is denoted by α^i and $X^{i+1} = X^i + \alpha^i S^i$; step 4) evaluate $f(X^{i+1})$ and $\nabla f(X^{i+1})$; step 5) set $Y^i = \nabla f(X^{i+1}) - \nabla f(x^i)$; step 6) $H^{i+1} = H^i + A^i + B^i$ where $A^i = \alpha^i S^i(S^i)^T/S^i\cdot Y^i$ and $B^i = -(H^iY^i)(H^iY^i)^T/Y^{iT}H^iY^i$; and step 7) $i = i + 1$ and return to step 1. The procedure is terminated when $\|S^i\|$ and $\alpha^i\|S^i\|$ are sufficiently small.

Conjugate Gradient Technique of Fletcher and Reeves

A necessary condition that $f(X)$ have a minimum at a point X is that $\nabla f(X) = 0$. Hence, by considering Eq. (2), we see that if the linear system $[\partial^2 f/\partial x^2]Y = B$ is solved for Y, where $B = [\partial^2 f/\partial x^2]X$, we will have located the minimum. The Fletcher-Reeves method solves this system for Y, although $[\partial^2 f/\partial x^2]$ is never evaluated by generating a sequence ∇f^i of gradient vectors of $f(X)$ which are $[\partial^2 f/\partial x^2]$ conjugate. That is, $\nabla f^{iT}[\partial^2 f/\partial x^2]\nabla f = \delta^{ij}a_j = 0$ if $i \neq j$ and a_j if $i = j$ with a_j constant.

This conjugate gradient method is an adaptation of work done by Hestenes and Stiefel.[9] Myers[10] has shown that for a quadratic function, the direction vectors generated by the Fletcher-Reeves method are scalar multiples of those generated by the Fletcher-Powell method, provided the initial step taken by each is in the direction of steepest descent. The algorithm is as follows: step 1) X^0 arbitrary, $P^0 = -\nabla f(X^0)$, $i = 0$, $j = 0$; step 2) X^{i+1} is the position of the minimum of $f(X)$ on the line through X^i in the direction P^i; step 3) evaluate $f(X^{i+1})$ and $\nabla f(X^{i+1})$; step 4) $\beta^i = \|\nabla f(X^{i+1})\|^2/\|\nabla f(X^i)\|^2$; step 5) $P^{i+1} = -\nabla f(X^{i+1}) + \beta^i P^i$, $j = j + 1$; and step 6) $i = i + 1$; if $j < n + 2$ go to step 2, otherwise set $X^0 = X^i$, $P^0 = -\nabla f(X^i)$, $j = 0$, and then go to step 2.
The iterations are terminated when $\|P^i\| < \epsilon$.

This algorithm also replaces the n dimensional problem with a sequence of one-dimensional minimizations. However, it has been the experience of many investigators that for this method the one-dimensional minimization does not have to be quite so accurate. It has also been found that the convergence is accelerated by reinitializing the search direction P^i to $+\nabla f$ every $n + 2$ cycles (recall step 6). This is because for a nonquadratic function this technique tends to generate nearly parallel search directions after $n + 2$ cycles. Notice also that this is the only method that does not compute the correction vector by multiplying the gradient by some positive definite matrix.

Goldstein-Price

This algorithm avoids the problem of finding a one-dimensional minimum, but we have the task of inverting a matrix and finding a lower value of $f(X)$ in the direction of the correction vector. Let $0 < \delta < \frac{1}{2}$ and $r > 0$ be given: step 1) compute the $n \times n$ matrix $Q(X^i)$ whose jth column is $[\nabla f(X^i + \theta_i I_j) - \nabla f(X^i)]/\theta_i$ where $\theta_0 = r$ and $\theta_j = r\|\Phi(X^{i-1})\|$, $i = 1, 2, \ldots$, where $\Phi(X^i)$ is defined as follows; step 2) if $k = 0$ or $Q(X^i)$ is singular or if $(\nabla f(X^i)^T Q^{-1}(X^i) f(X^i)) \leq 0$ set $\Phi(X^i) = Q^{-1}(X^i)\nabla f(X^i)$; step 3) now compute $\{f(X^i) - f[X^i - \nu\Phi(X^i)]\}/\nu[\nabla f(X^i)^T\Phi(X^i)] = g(X^i,\nu) \leftarrow \nu$ if $g(X^i,1) < \delta$, choose ν_i such that $\delta \leq g(X^i,\nu_i) \leq 1 - \delta$ otherwise set $\nu_i = 1$; and step 4) set $X^{i+1} = X^i - \nu_i\Phi(X^i)$.

The procedure is stopped when $\|\Phi(X^i)\| < \epsilon$. Notice that we are approximating $[\partial^2 f/\partial x^2]$ by the matrix $Q(X^i)$, which is found by differencing. If $Q(X^i)$ is singular or not "downhill," we are using the method of steepest descent (Curry[11]). The value of r should be chosen so that $Q(X^i)$ approximates $[\partial^2 f/\partial x^2]$ closely. The $\delta > 0$ implies $f(X^i)$, $i = 1, 2, \ldots$ is a decreasing sequence, and the closer to $\frac{1}{2}$ we choose δ, the more we force $f(X - \lambda\Phi)$ to approach its minimum with respect to λ.

Davidon's Second Method

Davidon recently published another method of minimizing a function.[6] An earlier version of this method was given in the AEC report containing his first method.

You will notice that in this algorithm both the problem of finding a one-dimensional minimum and that of inverting a matrix are avoided. However, for this algorithm at convergence we still have a good approximation for the inverse of the matrix of second partials at the minimum, and for a quadratic, Davidon showed this estimate to be exact.

Let V^0 be the initial positive definite estimate of the inverse of the matrix of second partials; V^i will be changed sequentially. Let α and β be positive real numbers satisfying $\alpha < 1 < \beta$. These numbers determine bounds on the change in V^i allowed from iteration to iteration. For a function which is nearly quadratic in nature, we choose β large and α near zero. If the function is highly nonquadratic we let α and β be near 1, so that we do not incorporate too much false information in V. The numbers α and β have the property that for every iteration and for all $U\epsilon E^n$, $\alpha U^T V^i U \leq U^T V^{i+1} U \leq \beta U^T V^i U$.

The algorithm is as follows: step 1) let $X^* = X^i - V^i\nabla f(X^i)$ and compute $f(X^*)$ and $\nabla f(X^*)$; step 2) define $r = V^i\nabla f(X^*)$ and $\rho = \nabla f(X^*)t_r$; if $\rho < \epsilon$ the final estimate for the location of the minimum is X^*; step 3) define $\nu = -[\nabla f(X^i)^T\cdot r]/\rho$ if $-\alpha/(1+\alpha) \leq \nu \leq \alpha/(1-\alpha)$, define $\lambda = \alpha$ if $-\beta/(\beta+1) \leq \nu \leq -\alpha/(1+\alpha)$, define $\lambda = -\nu/(\nu-1)$ if $-\beta/(\beta-1) \leq \nu \leq -\beta/(\beta+1)$, define $\lambda = \beta$; otherwise $\lambda = \nu/(\nu+1)$ and $V^{i+1} = V^i + (\lambda - 1)r(r^T/\rho)$; and step 4) if $f(X^i) \leq f(X^*)$ then $X^{i+1} = X^i$, $f(X^{i+1}) = f(X^i)$, $\nabla f(X^{i+1}) = \nabla f(X^i)$; if $f(X^*) < f(X^i)$ then let $X^{i+1} = X^*$, $f(X^{i+1}) = f(X^*)$, $\nabla f(X^{i+1}) = \nabla f(X^*)$, go to step 1.

Table 2 Comparison of algorithms on the second example

Method	Iterations	Function evaluations	Gradient evaluations	Value of function at convergence
Fletcher-Powell[a]	12	82	82	1.4×10^{-20}
Fletcher-Powell[b]	11	88	88	1.4×10^{-20}
Fletcher-Powell[c]	11	224	11	1.4×10^{-20}
Fletcher-Reeves	31	195	195	1.4×10^{-14}
Goldstein-Price	15	76	76	1.2×10^{-14}
Davidson's second	33	33	33	4.6×10^{-19}
Armstrong-Marquardt	12	12	12	2.8×10^{-18}

[a] Davidon's cubic interpolation used in one-dimensional minimization.
[b] Method of false position used in one-dimensional minimization.
[c] Method of golden sections and cubic fit used in one-dimensional minimization.

Armstrong-Marquardt Algorithm

This algorithm is applicable on a special class of functions, but they are frequently encountered so it is included in the discussion.

This class of functions is of the type $f[e(X)] = [e(X)^T \cdot Be(X)]/2$ where $e(X)$ is an m vector of functions of X, and B is a constant positive definite matrix. In order to derive the formula for the correction vector δx, we seek δx which minimizes

$$\Delta f[e(X)] = f\left(e(X) + \frac{\partial e(X)}{\partial x}\,\delta x\right) - f[e(X)] \cong f[e(X + \delta x)] - f[e(X)]$$

subject to the constraints that 1), $\|\delta x\| < \gamma$ and 2) $\Delta f[e(X)] < 0$. The correction vector δx is then found to be given by

$$\delta x = \{\partial e(X)^T/\partial x]B[\partial e(X)/\partial x] + \lambda I\}^{-1}[\partial e(X)^T/\partial x]\mathrm{Be}(X)$$

where $\|\delta x\| = \nu^2$, $\lambda > 0$.

In practice it was found that rather than working with constraints Eqs. (1) and (2) to compute λ, the $\lambda > 0$ was chosen so that Eq. (2) was satisfied (i.e., a decreasing sequence of function values was generated) and so that the matrix

$$[\{\partial e(X)^T/\partial x\}B\{\partial e(X)/\partial x\}] + \lambda I]$$

could be inverted.

Notice that if one computes the gradient of f with respect to X, then $\nabla f[e(X)] = [\partial e(X)^T/\partial x]Be(X)$. And clearly the matrix to be inverted is positive definite since $\lambda > 0$. So this correction vector is given by the product of a positive definite matrix times the gradient of f.

Also it was shown in Ref. (7) that as $\lambda \to \infty$, δx goes toward the direction of the gradient, that is, the steepest descent. And for $e(X)$, and n vector, as $\lambda \to 0$, it is seen that δx approaches the Newton-Raphson correction vector for finding $e(X) = 0$.

Functions to Be Minimized

In order to compare the relative convergence speed of the different methods, we first minimized a standard test function using each technique. The function which was minimized was Rosenbrock's parabolic valley function

$$f(x_1,x_2) = (x_2 - x_1^2)^2 + 0.01(x_1 - 1.0)^2$$

with starting values $(-1.2, +1.0)$. This function was chosen because it exhibits a long parabolic valley along $x_2 = x_1^2$.

The linear minimization necessary in the Fletcher-Powell method was carried out by the Johnson-Myers method[12] of golden sections and cubic fit on one run, and on the other the method of false position was used to find α such that $S^T\nabla f(x + \alpha S) = 0$. The linear minimizations of conjugate gradient were computed by the method of false position. The results are given in Table 1.

In order to illustrate the use of the techniques previously discussed, with respect to some of the optimal control problems in aerospace work, we chose a two-point boundary-value problem arising from the application of the maximum principle[13] to a simplified example. The example was a two-dimensional minimum time to intercept problem with constant thrust in free space. The equations of motion of this problem are

$$\dot{u} = T\cos\beta,\ \dot{v} = T\sin\beta,\ \dot{x} = u,\ \dot{y} = v \qquad (7)$$

where T is a constant. We wish to find $\beta = \beta(t)$ for the $t\epsilon(t_0,t_f)$ which minimizes

$$J[\beta] = \int_{t_0}^{t_f} dt$$

and satisfies the boundary conditions

$$x(t_0) = x_0 \qquad x(t_f) = \text{free}$$
$$y(t_0) = y_0 \qquad y(t_f) = y_f$$
$$v(t_0) = v_0 \qquad v(t_f) = v_f$$
$$u(t_0) = u_0 \qquad u(t_f) = u_f$$

The maximum principle then states that if an optimal control exists it must satisfy $\tan\beta = c_1 - c_2 t$, the linear tangent law. Using $\beta = \tan^{-1}(c_1 - c_2 t)$ in eqs. (7), we then estimate c_1, c_2, and t_f and integrate the system. Forming the function of three variables

$$f(c_1,c_2,\bar{t}_f) = [y(\bar{t}_f) - y_f]^2 + [v(\bar{t}_f) - v_f]^2 + [u(\bar{t}_f) - u_f]^2$$

we see that the triple $(c_1c_2\bar{t}_f)$ which minimizes f (that is, $f(c_1c_2\bar{t}_f) = 0$) solves the boundary-value problem and satisfies the necessary condition for optimality.

The partials of f with respect to c_1 and c_2 were calculated by taking the partials of eqs. (7) with respect to c_1 and c_2 and integrating the resulting equations. This interchanging of limits is justified by the smoothness of the functions involved.[14] The partial of f with respect to t_f was calculated by evaluating Eq. (7) at t_f.

The linear minimizations for the Fletcher-Powell technique were performed using; 1) the method of false position, 2) golden sections and cubic fit, and 3) the cubic interpolation of Davidon outlined in Eq. 5. The linear minimizations for the conjugate gradient were computed as in Eq. (5) also. The results in terms of number of iterations and number of function and gradient evaluations are given in Table 2.

Comparison of Methods

For the majority of practical problems, in which the solution depends upon minimizing an unconstrained function of several variables, the calculation of the function takes much more computer time than do the calculations in any of the minimization routines. It is for this reason that the basis of comparison of these algorithms is upon the number of function evaluations needed to achieve a minimum rather than the number of minimization iterations. Notice that each of the algorithms generates a monotone decreasing sequence of function values, and they are generally easy to program. So all the algorithms satisfy two of the three criteria set forth in the introduction.

On the basis of the number of function evaluations necessary to achieve convergence for functions of the type $f(X) = [e(X)^T Be(X)]/2$ with B a positive definite matrix of constants, it seems that the Armstrong-Marquardt combined steepest descent and Newton-Raphson method is best. This is seen in Table II and from other problems solved by these techniques.

For the general problem, the Fletcher-Powell and Fletcher-Reeves techniques are the easiest to implement if an effective and efficient one-dimensional minimization scheme is available. This is because, other than convergence criterion, there are no algorithm dependent parameters to estimate as in the others. The Fletcher-Powell technique has shown a faster convergence rate than the conjugate gradient. It also requires a more accurate linear search. Also at convergence the Fletcher-Powell method yields an estimate of the inverse of the matrix of second partial derivatives.

The Goldstein-Price Algorithm has demonstrated the slowest convergence initially but near the minimum it has shown the fastest. However, the Q matrix which approximates the matrix of second partials is computed by one-sided differencing of the gradient. Since in each iteration the same perturbation increment is used for all variables, difficulties can arise if the gradient is much more sensitive to a change in one variable than another. In this case Q may not be a good approximation to $[\partial^2 f/\partial x^2]$ or Q may be singular or cause the step to be uphill. This forces the Goldstein-Price algorithm to use steepest descent[11] so the convergence slows. Also it takes N gradient evaluations to compute Q. Thus, if N the number of variables is greater than say 20, the Fletcher-Powell method with any of the linear-minimization techniques mentioned would be faster.

The second method of Davidon has shown the fastest overall convergence rate in terms of function evaluations. Also, it should be noted that both the problems of inverting a matrix and finding a one-dimensional minimum are avoided in this method. However, the choice of α, β, and ϵ is very important. It was found that the fastest convergence occurred when

$$\alpha\epsilon[0.001, 0.5],\ \beta\epsilon[3.0, 100.0],\ \text{and}\ \epsilon\epsilon[10^{-25}, 10^{-10}]$$

Also, on some problems the gradient at the initial point may be so large that if $V^0 = I_n$, the correction vector is unrealistically large. Problems of this nature can be avoided by using $V^0 = tI_n$ where $t > 0$ is small enough to cut the step size to an acceptable level.

References

[1] Steward, G. W., III, "A Modification of Davidon's Minimization Method to Accept Difference Approximations of Derivatives," *Journal of ACM*, Vol. 14, No. 1, Jan. 1967, pp. 72–83.

[2] Fletcher, R. and Powell, M. J. S., "A Rapidly Convergent Descent Method for Minimization," *Computer Journal*, Vol. 6, 1963, pp. 163–168.

[3] Davidon, W. C., "Variable Metric Method for Minimization," ANL-5990, Rev., 1959, Atomic Energy Commission Research Development.

[4] Goldstein, A. A. and Price, J. F., "An Effective Algorithm for Minimization," Mathematical Note 483, AD 643 973, Sept. 1966, Boeing Scientific Research Lab.,

[5] Fletcher, R. and Reeves, C. M., "Function Minimization by Conjugate Gradients," *Computer Journal*, Vol. 7, 1964, pp. 149–154.

[6] Davidon, W. C., "Variance Algorithm for Minimization," *Computer Journal*, Vol. 10, No. 4, Feb. 1968, pp. 406–411.

[7] Armstrong, E. S., "A Combined Newton-Raphson and Gradient Parameter Correction Technique for Solution of Optimal-Control Problems," TR R-293, 1968, NASA.

[8] Marquardt, D. W., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM Journal*, Vol. 11, No. 2, June 1963, pp. 431–441.

[9] Hestenes, M. R. and Steifel, E., "Methods of Conjugate Gradients for Solving Linear Systems," *Journal of Research of the National Bureau of Standards*, Vol. 49, No. 6, Dec. 1952, pp. 409–436.

[10] Myers, G. E., "Properties of the Conjugate Gradient and Davidon Methods," CR-92127, 1968, NASA.

[11] Curry, H. B., "The Method of Steepest Descent for Nonlinear Minimization Problems," *Quarterly of Applied Mathematics*, Vol. II, Oct. 1944.

[12] Guilfoyl, G., Johnson, I., and Wheatley, P., "One-Dimensional Search Combining: Golden Section and Cubic Techniques," CR-65994, Jan. 1967, NASA.

[13] Pontryagin, L. S. et al., *The Mathematical Theory of Optimal Processes*, Wiley, New York, 1962.

[14] Widder, D. V., *Advanced Calculus*, Prentice-Hall, Englewood Cliffs, N. J., 1961.

[15] Ralston, A., *A First Course in Numerical Analysis*, McGraw-Hill, New York, 1965.

[16] Kelley, H. J. et al., "An Accelerated Gradient Method for Parameter Optimization With Nonlinear Constraints," *Journal of Astronautical Sciences*, Vol. 13, No. 4, July-Aug. 1966, pp. 166–169.