NASA TECHNICAL REPORT

NASA TR R-394

NASA TR R-394

DAVIDON-BROYDEN RANK-ONE MINIMIZATION METHODS IN HILBERT SPACE WITH APPLICATION TO OPTIMAL CONTROL PROBLEMS

by Terry A. Straeter
Langley Research Center
Hampton, Va. 23365

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • DECEMBER 1972

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TR R-394		

4. Title and Subtitle	5. Report Date
DAVIDON-BROYDEN RANK-ONE MINIMIZATION METHODS IN HILBERT SPACE WITH APPLICATION TO OPTIMAL CONTROL PROBLEMS	December 1972
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Terry A. Straeter	L-8576
	10. Work Unit No.
9. Performing Organization Name and Address	501-06-01-09
NASA Langley Research Center Hampton, Va. 23365	11. Contract or Grant No.
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Technical Report
National Aeronautics and Space Administration Washington, D.C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes The material presented herein includes information from a thesis entitled "On the Extension of the Davidon-Broyden Class of Rank-One, Quasi-Newton Minimization Methods to an Infinite Dimensional Hilbert Space With Applications to Optimal Control Problems" by Terry Anthony Straeter submitted to North Carolina State University, Raleigh, North Carolina, 1971, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Applied Mathematics.

16. Abstract

The Davidon-Broyden class of rank-one, quasi-Newton minimization methods is extended from Euclidean spaces to infinite dimensional real Hilbert spaces. The members of this class of minimization methods are distinguished by the manner in which a particular parameter (the step size) is chosen at each iteration. For several techniques of choosing the step size, conditions are found which assure convergence of the associated iterates to the location of the minimum of a positive definite quadratic functional. For those techniques, convergence is achieved without the problem associated with many other first-order minimization methods, namely, the computation of a one-dimensional minimum at each iteration. The application of this class of minimization methods for the direct computation of the solution of an optimal control problem is outlined. The performance of various members of the class are compared by solving a sample optimal control problem. Finally, the sample problem is solved by other known gradient methods, and the results are compared with those obtained with the rank-one quasi-Newton methods.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Optimal control Minimization Direct methods Hilbert space Functional analysis	Unclassified – Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	31	$3.00

*For sale by the National Technical Information Service, Springfield, Virginia 22151

DAVIDON-BROYDEN RANK-ONE MINIMIZATION METHODS IN HILBERT SPACE WITH APPLICATION TO OPTIMAL CONTROL PROBLEMS *

By Terry A. Straeter
Langley Research Center

SUMMARY

The Davidon-Broyden class of rank-one quasi-Newton minimization methods is extended from Euclidean spaces to infinite dimensional real Hilbert spaces. The members of this class of minimization methods are distinguished by the manner in which a particular parameter (the step size) is chosen at each iteration. For several techniques of choosing the step size, conditions are found which assure convergence of the associated iterates to the location of the minimum of a positive definite quadratic functional. For those techniques, convergence is achieved without the problem associated with many other first-order minimization methods, namely, the computation of a one-dimensional minimum at each iteration. The application of this class of minimization methods for the direct computation of the solution of an optimal control problem is outlined. The performance of various members of the class are compared by solving a sample optimal control problem. Finally, the sample problem is solved by other known gradient methods and the results are compared with those obtained with the rank-one quasi-Newton methods.

INTRODUCTION

In the past few years the problem of finding the location of the minimum value of a real-valued function of n real variables by numerical methods has been the subject of a great deal of research. (See refs. 1 to 4.) Several iterative procedures have been developed to solve the problem. Much of the work has been directed toward developing algorithms which use the function value and its gradient to locate the minimum by iteration. This type of algorithm is usually referred to as a gradient method. Historically, the method of steepest descent was the first such method. In order to accelerate conver-

*The material presented herein includes information from a thesis entitled "On the Extension of the Davidon-Broyden Class of Rank-One, Quasi-Newton Minimization Methods to an Infinite Dimensional Hilbert Space With Applications to Optimal Control Problems" by Terry Anthony Straeter submitted to North Carolina State University, Raleigh, North Carolina, 1971, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Applied Mathematics.

gence, the method of conjugate gradients, developed by Hestenes and Stiefel (ref. 5), was applied to the minimization problem by Fletcher and Reeves (ref. 4). Later various versions of it were extended to real Hilbert spaces (refs. 6 and 7) and applied to optimal control problems (refs. 8 and 9). Other first-order methods were developed which were inspired by Newton's second-order method. In 1959 Davidon (ref. 2) proposed two effective techniques for solving the problem. The first method, given in the main body of his report, was modified in 1963 by Fletcher and Powell (ref. 3). This algorithm is referred to as the DFP method. They established that for any differentiable real-valued function, the method is stable; that is, the value of the function is monotonically decreasing with each iteration. Moreover, they showed that for a real-valued quadratic function of n variables, the DFP algorithm converges in a finite number of steps. In fact, at most $n + 1$ steps are needed. In 1968 Horwitz and Sarachik (ref. 10) extended the DFP method from an n-dimensional Euclidean vector space to an infinite-dimensional real Hilbert space and established convergence of the iterates when the functional to be minimized is quadratic. In 1970 Tokumaru, Adachi, and Goto (ref. 11) also extended the DFP algorithm to an infinite-dimensional real Hilbert space and gave a comparison of the DFP method, the steepest descent method, and the conjugate gradient method on some sample optimal control problems.

The second method due to Davidon, which he called a rank-one variance method, was outlined in the appendix to reference 2. In reference 12 and again in reference 13, he published modifications of the second method and established conditions insuring its convergence to the minimum of a quadratic function of n variables in a finite number of steps, at most $n + 1$, and insuring the stability of the method. In 1965 Broyden (ref. 1) proposed a family of methods, which he called quasi-Newton, based on an arbitrary parameter α (the step size). If $\alpha = 1$, then under certain conditions, Broyden's method and the second Davidon method are the same. In 1969 Goldfarb (ref. 14) established convergence of the iterates of a rank-one algorithm for a class of real functions of n variables when α is chosen by means of a linear minimization technique (that is, a one-dimensional search). He called the algorithm with the step size chosen by a one-dimensional minimization "an optimal variance algorithm."

The purpose of this paper is to extend the Davidon-Broyden family of algorithms to an infinite-dimensional real Hilbert space, to establish conditions guaranteeing convergence of the iterates for various algorithms in the family, and to apply the family of algorithms to optimal control problems. Also presented herein is a comparison of the rank-one methods with other first-order algorithms on a sample optimal control problem.

SYMBOLS

A	self-adjoint, positive, linear operator from H to H
A^{-1}	inverse of A
$B^{(n)}, V^{(n)}$	sequence of linear operators
b	fixed element of H
C	linear operator from H to H
f	function from R^{n+m+1} into R^n
f_u	$n \times m$ matrix with elements $\left(\frac{\partial f_i}{\partial u_j}\right)$ where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$
f_x	$n \times n$ matrix with elements $\left(\frac{\partial f_i}{\partial x_j}\right)$ where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$
g	gradient of J
g_n, g^*	gradient J at x_n and x^*, respectively
H	real Hilbert space
h	element of H
I	identity
i, j, k, N, n, q, r	integers
J	functional defined on H
$J_0, M, m, M_0, m_0, \tilde{J}_m, K$	real numbers

L	function from R^{n+m+1} into R
$L_x = \left(\frac{\partial L}{\partial x_1}, \frac{\partial L}{\partial x_2}, \ldots, \frac{\partial L}{\partial x_n}\right)$	
$\mathcal{L}_r^2(t_0,t_1)$	square integrable functions from (t_0,t_1) into R^r
$M(n), N(n)$	set of integers
$P(t), R(t), G(t), B(t)$	matrix-valued function of t
p_i	elements of H
R	set of all real numbers
r_n	nth residual vector, element of H, defined by equations (10)
S, M, S'_x, S'_{x_0}	subspaces of H
s_n	direction of nth step, element of H
t	independent variable
t_0, t_1	initial and final values of the independent variable
U_x	linear operator from H to H
u	control variable
x, u, x_0, x^*	elements of H
$\tilde{x}$	element of H at which J is minimized
x_1, x_2, x	state variables
x_n	nth iterate, element of H
y_n	element of H defined by $g^* - g_n$

$\alpha,\beta,\beta_i,\lambda,\mu$	scalars
α_n	step size, a real number
δ_u	change in u
ϵ_n	scalar, $1 - \alpha_n$
$\epsilon_1(h)$	functional from H to R
θ	zero of H
$\lambda,\lambda_1,\lambda_2$	adjoint variables
ρ_i	scalar defined by equation (14)
σ_n	defined by $x_{n+1} - x_n$
τ_n	scalars defined by equations (11)
∇_u	gradient with respect to u
$(\cdot,\cdot)$	inner product on H
$\|\| \cdot \|\|$	norm on H defined as $(\cdot,\cdot)^{1/2}$, and norm of operator A defined by $\|\|A\|\| = \inf\{K:\|\|Ax\|\| \leq K\|\|x\|\|\}$

Abbreviations:

conv	convex hull
dim(H)	dimension of H
inf	infimum (lower bound)
sup	supremum (upper bound)

Dots over symbols denote derivatives with respect to t.

PROBLEM FORMULATION FOR QUADRATIC CASE

Consider the problem of finding the element $\tilde{x} \in H$ which minimizes the quadratic functional $J:H \to R$ given by

$$J(x) = J_0 + (x,b) + \frac{1}{2}(x,Ax) \tag{1}$$

where $J_0 \in R$, the real numbers, and b is a fixed element in H, a real Hilbert space with inner product $(\cdot,\cdot)$ and zero θ.

In equation (1) $A:H \to H$ is a linear self-adjoint operator; thus,

$$m(x,x) \leqq (x,Ax) \leqq M(x,x) \tag{2}$$

where

$$M = \sup_{x \neq \theta} \frac{(x,Ax)}{(x,x)}$$

$$m = \inf_{x \neq \theta} \frac{(x,Ax)}{(x,x)}$$

and where it is assumed that $0 < m \leqq M$. Hence, $||A|| = M$ (ref. 15) where

$$||A|| = \inf\{K: ||Ax|| \leqq K||x||\}$$

Since $m > 0$, A^{-1} exists (ref. 16) and A^{-1} is also a self-adjoint operator. Moreover,

$$\frac{1}{M}(x,x) \leqq \left(x,A^{-1}x\right) \leqq \frac{1}{m}(x,x) \tag{3}$$

A functional $J:H \to R$ is said to be differentiable at x if there exists a bounded linear functional $U_x:H \to R$ such that for $h \in H$

$$J(x + h) - J(x) = U_x(h) + \epsilon_1(h) \tag{4}$$

where $\dfrac{\epsilon_1(h)}{(h,h)^{1/2}} \to 0$, as $(h,h)^{1/2} \to 0$ (Fréchet differential). If such a functional U_x exists, then it is unique. (See ref. 17.) Moreover, by the Riesz representation theorem (ref. 18), for each $x \in H$ there exists a vector $g \in H$ such that $(g,h) = U_x(h)$ for all $h \in H$ and g is given by

$$(g,h) = \left.\frac{dJ(x + th)}{dt}\right|_{t=0}$$

The vector g is called the gradient of the functional J.

For the quadratic functional defined in equation (1), the gradient is given by

$$g(x) = Ax + b \tag{5}$$

A well-known necessary and sufficient condition that $\tilde{x}$ minimize $J(x)$ as given by equations (1) and (2) is that $g(\tilde{x}) = \theta$ where θ denotes the zero element of H. (See ref. 15.)

Hence, if $\tilde{x}$ denotes the location of the minimum of the quadratic functional J given by equation (1), then

$$\tilde{x} = -A^{-1}b \tag{6}$$

Moreover, if $x,h \in H$ are such that $x + h = \tilde{x}$, then

$$h = -A^{-1}g(x) \tag{7}$$

Of course, the equation $h = -A^{-1}g(x)$ is the basis for the well-known Newton-Raphson method for solving the operator equation $g(x) = \theta$ on a real Hilbert space. (See ref. 15.) As a final preliminary note, recall that if A and B are positive self-adjoint linear operators with $(x,Ax) \geqq (x,Bx)$ for all $x \in H$, it is said that $A \geqq B$. (See ref. 18.)

DAVIDON-BROYDEN ALGORITHMS

Let $J:H \to R$ be a differentiable functional with gradient g. Let $x_0 \in H$ be the initial estimate of the location of the minimum of J, and let $V^{(0)}$ be a self-adjoint linear operator from H onto H. Moreover, let $M_0 \geqq m_0 > 0$ be such that for all x, $m_0(x,x) \leqq (x, V^{(0)}x) \leqq M_0(x,x)$, that is, $V^{(0)}$ is strongly positive. (See ref. 18.) If J, the functional to be minimized, happens to be quadratic, then $V^{(0)}$ can be viewed as an estimate of A^{-1}. The quantities $J(x_0)$ and $g(x_0)$ are computed and the first ($n = 0$) and successive iterations are obtained as follows:

(1) Let

$$x^* = x_n - \alpha_n V^{(n)} g_n \tag{8}$$

where g_n denotes $g(x_n)$ and α_n, the step size, is a scalar, the choice of which is discussed later. Let

$$s_n = -V^{(n)}g_n \tag{9}$$

and compute $J(x^*)$ and $g(x^*)$ denoted by g^*; if $(g^*,g^*) = 0$, a necessary condition for x^* to be the location of the minimum, the computation is stopped. If J is a quadratic functional as in equations (1) and (2) and $g^* = \theta$, then x^* is the location of the minimum. (See ref. 15.)

(2) Compute the vector

$$\left.\begin{aligned} r_n &= V^{(n)}g^* - V^{(n)}g_n + \alpha_n V^{(n)}g_n \\ r_n &= V^{(n)}\left[g^* - (1 - \alpha_n)g_n\right] \\ r_n &= V^{(n)}y_n - \alpha_n s_n \end{aligned}\right\} \tag{10}$$

where

$$y_n = g^* - g_n$$

The vector r_n is called the residual vector for reasons explained in a subsequent section. If $r_n = \theta$ and $\alpha_n \neq 1$, set $\alpha_n = 1$ and return to step (1). If $r_n = \theta$ and $\alpha_n = 1$, then set $\tau_n = 0$ and $V^{(n+1)} = V^{(n)}$ and go to step (5).

(3) Define the scalar τ_n as

$$\tau_n = \begin{cases} -(y_n, r_n)^{-1} & \left((y_n, r_n) \neq 0\right) \\ 0 & \left((y_n, r_n) = 0\right) \end{cases} \tag{11}$$

(4) Let

$$V^{(n+1)} = V^{(n)} + \tau_n B^{(n)} \tag{12}$$

where $B^{(n)}: H \to H$ is defined such that for all $x \in H$

$$B^{(n)}x = (x, r_n)r_n \tag{13}$$

(5) If $J(x^*) < J(x_n)$, let $x_{n+1} = x^*$ and, consequently, $J(x_{n+1}) = J(x^*)$ and $g_{n+1} = g^*$. Otherwise, if $\tau_n \neq 0$, let $x_{n+1} = x_n$ so that $J(x_{n+1}) = J(x_n)$ and $g_{n+1} = g_n$. If $\tau_n = 0$, then return to step (1) and choose α_n so that $J(x^*) < J(x_n)$. Set $n = n + 1$ and go to step (1).

Figure 1 gives a two-dimensional illustration of the behavior of the algorithm. The figure depicts the level curves of a hypothetical function J, the iterates generated by the algorithm, the negative gradient at each iterate, and the direction of the step generated by equation (9) at each iterate. Notice when $V^{(0)}$ is the identity, then the first direction is the negative gradient. Also note that the illustration shows that $x_1 = x_2$; that is, $J(x^*)$ for the first iteration is greater than $J(x_1)$. Hence by step (5), $x_2 = x_1$. However, the computation of g^* is used to compute $V^{(2)}$ which determines the choice of s_2. The point x_3 is then such that $J(x_3)$ shows a substantial improvement over $J(x_2)$. From x_3, the iterates would continue until the necessary conditions for a minimum are satisfied.

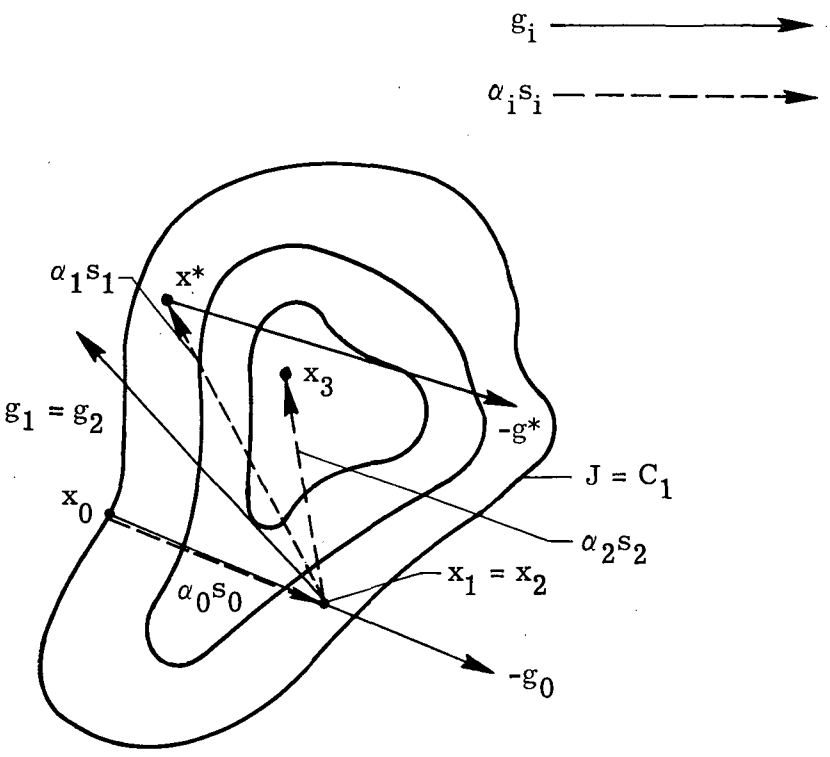

Figure 1.- Example of progress of algorithm.

Notice that the operator $V^{(n+1)}: H \to H$ given in equation (12) can be expressed as

$$V^{(n+1)}x = V^{(0)}x + \sum_{i \in M(n)} \rho_i (x, p_i)\, p_i \tag{14}$$

where $x \in H$ and $M(n) = \{i \leq n: (y_i, r_i) \neq 0\}$ and

$$\rho_i = -\operatorname{sgn} \frac{1}{(y_i, r_i)} \qquad (i \in M(n))$$

and

$$p_i = \frac{r_i}{\left[\left|(y_i, r_i)\right|\right]^{1/2}} \qquad (i \in M(n))$$

This expression for $V^{(n+1)}$ is computationally more tractable than equation (12) for those problems for which equation (12) would necessitate storing $V^{(n)}$ as a large rectangular array as in optimal control problems. However, equation (14) implies that only the representation of p_i needs to be stored with each iteration.

The elements of the class of algorithms outlined are distinguished by the manner in which the parameter α_n is chosen with each iteration. Davidon (ref. 2), Broyden (ref. 1), and Goldfarb (ref. 14) proposed techniques for choosing α_n in the finite-dimensional case. For Davidon's rank-one variance algorithm, $\alpha_n = 1$ for all n; however, the scalar τ_n given by equations (11) is chosen so that certain inequality constraints are satisfied. These constraints insure that Davidon's $V^{(n)}$ remain positive definite. Goldfarb's optimal variance algorithm required that α_n be chosen so that $J(x_n + \alpha s_n)$ be minimized with respect to α. The Broyden quasi-Newton method requires only that α_n be chosen so that $(V^{(n)})^{-1}$ exists and $(y_n, r_n) \neq 0$. It is shown subsequently that for certain choices of $V^{(0)}$, both Davidon's and Goldfarb's methods of choosing α_n satisfy Broyden's criteria.

THEOREMS INDEPENDENT OF CHOICE OF α_n

It is assumed that the functional to be minimized is quadratic as defined by equations (1) and (2). However, the first three theorems of this section are proved without using this fact. Hence, the results of these theorems hold if J is any differentiable function.

Theorem 1: $B^{(n)}$ as given in equation (13) is a linear self-adjoint, nonnegative, operator for all n, for any choice of α_n (nonnegative, that is, for all $x \in H$ $\left(x, B^{(n)}x\right) \geq 0$).

Proof: Clearly, $B^{(n)}$ is linear by the linearity of the inner product. If $x \in H$, then

$$\left(x, B^{(n)}x\right) = \left(x, (x, r_n)r_n\right) = (x, r_n)^2 \geq 0$$

and if $x, y \in H$, then

$$\left(x, B^{(n)}y\right) = \left(x, (y, r_n)r_n\right) = (y, r_n)(x, r_n) = \left(y, (x, r_n)r_n\right) = \left(y, B^{(n)}x\right)$$

Theorem 2: $V^{(n)}$ is self-adjoint for all n, for any choice of α_n.

Proof: $V^{(n)} = V^{(0)} + \sum_{i=0}^{n-1} \tau_i B^{(i)}$ by equation (14) and $V^{(0)}$ is self-adjoint by definition. By this theorem, the $B^{(i)}$ terms are self-adjoint and the finite sum of self-adjoint operators is self-adjoint.

Theorem 3 (Basic theorem): If $\tau_n \neq 0$ or $r_n = \theta$, then $V^{(n+1)}(g^* - g_n) = x^* - x_n$; that is, $V^{(n+1)}y_n = \alpha_n s_n$.

Proof: If $r_n = \theta$, then $V^{(n)}y_n = \alpha_n s_n$ by definition. Also since $\tau_n = 0$ by equations (11) and (12), $V^{(n+1)} = V^{(n)}$; thus, $V^{(n+1)}y_n = \alpha_n s_n$. Otherwise, consider

$$\begin{aligned} V^{(n+1)}y_n - \alpha_n s_n &= V^{(n)}y_n + \tau_n(y_n, r_n)r_n - \alpha_n s_n \\ &= V^{(n)}y_n - \frac{(y_n, r_n)}{(y_n, r_n)} r_n - \alpha_n s_n \\ &= V^{(n)}y_n - \alpha_n s_n - r_n = \theta \end{aligned}$$

by definition of r_n.

The rank-one algorithms are Newtonlike in that the change in x at each iteration is given by a linear operator upon the gradient at x_n. The name quasi-Newton was given to these algorithms by Broyden (ref. 1) because of the property of $V^{(n+1)}$ given in theorem 3. Notice that for a quadratic functional such as equation (1), A^{-1} also has this property, that is, $A^{-1}(g^* - g_n) = x^* - x_n$ or

$$A^{-1}y_n = \alpha_n s_n \tag{15}$$

Therefore, $V^{(n+1)}$ and A^{-1} agree on the vector $g^* - g_n$.

The next theorem and its corollary state that because of the manner in which $B^{(n)}$ is defined, if A^{-1} and $V^{(n)}$ agree on a vector, then A^{-1} and $V^{(n+1)}$ also do. That is, information about the nature of A^{-1} which $V^{(n)}$ contains will be retained by $V^{(n+1)}$.

Theorem 4: If $u \in H$ is such that $A^{-1}u = V^{(n)}u$ and $C:H \to H$ is a linear operator such that $C = V^{(n)} + \mu B^{(n)}$ for some real μ, then $A^{-1}u = Cu$.

Proof: Since $A^{-1}(g^* - g_n) = x^* - x_n$ and $x^* = x_n - \alpha_n V^{(n)} g_n$ by definition,

$$x^* - x_n = -\alpha_n V^{(n)} g_n = A^{-1}(g^* - g_n)$$

and

$$r_n = V^{(n)}(g^* - g_n) + \alpha_n V^{(n)} g_n$$

Therefore, $r_n = V^{(n)}(g^* - g_n) - A^{-1}(g^* - g_n)$ and, hence,

$$r_n = \left(V^{(n)} - A^{-1}\right)(g^* - g_n) \tag{16}$$

Since $A^{-1}u = V^{(n)}u$,

$$\left(V^{(n)} - A^{-1}\right)u = \theta \tag{17}$$

Therefore

$$(r_n, u) = \left(\left(V^{(n)} - A^{-1}\right)(g^* - g_n), u\right) = \left(g^* - g_n, \left(V^{(n)} - A^{-1}\right)u\right) = (g^* - g_n, \theta) = 0$$

by theorem 2 and equations (16) and (17). Hence, the hypothesis $\left(C - A^{-1}\right)u = \mu B^{(n)}u$ implies that

$$\mu B^{(n)}u = \mu(u, r_n) r_n = \mu \cdot 0 \cdot r_n = \theta$$

Therefore, $Cu = A^{-1}u$.

Since $V^{(n+1)} = V^{(n)} + \tau_n B^{(n)}$, the following corollaries are obtained.

Corollary 1: If $V^{(n)}u = A^{-1}u$ for some $u \in H$, then $V^{(n+1)}u = A^{-1}u$.

Corollary 2 (Fundamental property of $V^{(n)}$): $V^{(n)}y_i = \alpha_i s_i = A^{-1}y_i$ for all $i < n$ if $\tau_j \neq 0$ for $j = 0, 1, \ldots, n$.

Proof: Note as previously observed (eq. (15)) since J is quadratic, $A^{-1}y_i = \alpha_i s_i$. From theorem 3,

$$V^{(1)}y_0 = \alpha_0 s_0$$

Assume that $V^{(n)}y_i = \alpha_i s_i$ for all $i < n$. Consider $V^{(n+1)}y_i$ for $i = n$. Since $\tau_n \neq 0$ by theorem 3, $V^{(n+1)}y_n = \alpha_n s_n$. Otherwise, for $i < n$, A^{-1} and $V^{(n)}$ agree on y_i. Corollary 1 implies that $V^{(n+1)}y_i = \alpha_i s_i$. Thus the corollary follows by mathematical induction.

Corollary 2 is most useful in later convergence arguments and, hence, it has been named "the fundamental property of $V^{(n)}$."

By equation (15) and theorem 1, Goldfarb (ref. 14) observed that equation (12) can be expressed as

$$V^{(n+1)} = V^{(n)} - \frac{B^{(n)}}{\left(\left(V^{(n)} - A^{-1}\right)y_n, y_n\right)}$$

if $\tau_n \neq 0$; otherwise, $V^{(n+1)} = V^{(n)}$ which yields the following theorem.

Theorem 5: If $V^{(0)} \geqq A^{-1}$, then $V^{(n)} \geqq A^{-1}$ for all n and similarly, if $V^{(0)} \leqq A^{-1}$, then $V^{(n)} \leqq A^{-1}$ for all n.

Proof: Proof of this theorem is by mathematical induction. Assume that $V^{(n)} \geqq A^{-1}$. If $V^{(n+1)} = V^{(n)}$, that is, $\tau_n = 0$, the result is trivial. Otherwise, by the preceding equation

$$\left(x, \left(V^{(n+1)} - A^{-1}\right)x\right) = \left(x, \left(V^{(n)} - A^{-1}\right)x\right) - \frac{(x, r_n)^2}{\left(y_n, \left(V^{(n)} - A^{-1}\right)y_n\right)}$$

and by equations (10) and (15), $r_n = \left(V^{(n)} - A^{-1}\right)y_n$. The inductive hypothesis implies that $V^{(n)} - A^{-1}$ is a positive operator. Hence, by the Cauchy-Bunyakovski-Scharwz inequality (refs. 18 and 19),

$$\left(x, \left(V^{(n)} - A^{-1}\right)x\right) - \frac{\left(x, \left(V^{(n)} - A^{-1}\right)y_n\right)^2}{\left(y_n, \left(V^{(n)} - A^{-1}\right)y_n\right)} \geqq 0$$

The second part of the theorem is obtained by merely considering $\left(x, \left(A^{-1} - V^{(n+1)}\right)x\right)$ instead.

The following theorem whose finite dimensional analog is due to Goldfarb (ref. 14) gives a condition under which the $V^{(n)}$ operators form a monotone sequence of self-adjoint bounded linear operators.

Theorem 6: If $V^{(0)} \geq A^{-1}$, then $A^{-1} \leq V^{(n)} \leq \ldots \leq V^{(0)}$ for all n. Similarly, if $V^{(0)} \leq A^{-1}$, then $A^{-1} \geq V^{(n)} \geq \ldots \geq V^{(0)}$ for all n.

Proof: By theorem 5, if $V^{(0)} \geq A^{-1}$, then $V^{(n)} \geq A^{-1}$ for all n. If $V^{(n+1)} = V^{(n)}$, that is, $\tau_n = 0$, then the assertion is obvious. Otherwise,

$$\left(x, \left(V^{(n+1)} - V^{(n)}\right)x\right) = -\frac{\left(x, B^{(n)}x\right)}{\left(y_n, \left(V^{(n)} - A^{-1}\right)y_n\right)} \leq 0$$

by equation (12). The inequality holds since theorem 1 gives $\left(x, B^{(n)}x\right) \geq 0$ and from theorem 5, $V^{(n)} - A^{-1} \geq 0$. The second part of the theorem follows by considering $V^{(n)} - V^{(n+1)}$ instead.

Corollary 3: If $V^{(0)} \leq A^{-1}$ or $V^{(0)} \geq A^{-1}$, then the $V^{(n)}$ operators form a monotone sequence of strongly positive self-adjoint linear operators bounded by $V^{(0)}$ and A^{-1}. Moreover, there exists a strongly positive self-adjoint operator V such that $\lim_{n\to\infty} ||V^{(n)}x - Vx|| = 0$ for each $x \in H$.

Proof: The $V^{(n)}$ operators form a bounded monotone sequence of strongly positive, self-adjoint operators by theorems 2 and 6; that is, if $V^{(0)} \leq A^{-1}$, then $V^{(0)} \leq V^{(1)} \leq V^{(2)} \leq \ldots \leq V^{(n)} \leq \ldots A^{-1}$. This relationship implies the existence of a strongly positive self-adjoint linear operator V such that $V^{(n)}$ converges to V pointwise (ref. 18) and similarly, for $V^{(0)} \geq A^{-1}$.

Theorem 7: If $\tau_n \neq 0$ for all n and if S is the space spanned by $\{y_n\}$, then $\lim_{n\to\infty} ||V^{(n)}x = A^{-1}x||$ for all $x \in S$ independent of the choice of the α_n. (By the space spanned by a set M is meant the smallest linear manifold containing M.)

Proof: Since $x \in S$, this relationship means that x can be written as a finite linear combination of y_n vectors. Therefore there exist $\beta_i \in R$ and some k such that

$$x = \sum_{i=1}^{k} \beta_i y_i$$

Let $n > k$, then

$$V^{(n)}x = \sum_{i=1}^{k} \beta_i V^{(n)} y_i$$

$$= \sum_{i=1}^{k} \beta_i A^{-1} y_i \qquad \text{(Corollary 2)}$$

$$= A^{-1} \sum_{i=1}^{k} \beta_i y_i$$

$$= A^{-1} x$$

Theorem 8: If $V^{(0)} \leqq A^{-1}$ or $V^{(0)} \geqq A^{-1}$ and $\tau_n \neq 0$ for all n and the y_i form a Schauder basis (ref. 20) for H, then $V^{(n)} \to A^{-1}$ pointwise independent of the choice of α_n.

Proof: For any $x \in H$ there exist $\beta_i \in R$ such that

$$x = \sum_{i=0}^{\infty} \beta_i y_i \tag{18}$$

since $\{y_i\}$ form a Schauder basis for H. Consider

$$\left\| A^{-1}x - V^{(n)}x \right\| = \left\| \left(A^{-1} - V^{(n)} \right) x \right\|$$

$$= \left\| \left(A^{-1} - V^{(n)} \right) \sum_{i=0}^{\infty} \beta_i y_i \right\| \leqq \left\| \left(A^{-1} - V^{(n)} \right) \sum_{i=0}^{n-1} \beta_i y_i \right\| + \left\| \left(A^{-1} - V^{(n)} \right) \sum_{i=n}^{\infty} \beta_i y_i \right\|$$

By corollary 2, $\left(A^{-1} - V^{(n)} \right) \sum_{i=0}^{n-1} \beta_i y_i = \theta$. Since $V^{(0)} \geqq A^{-1}$ or $V^{(0)} \leqq A^{-1}$ by theorem 6 and its corollary, it must be that $\left\| V^{(n)} \right\| \leqq \left\| A^{-1} \right\|$ or $\leqq \left\| V^{(0)} \right\|$. Therefore $\left\| A^{-1} - V^{(n)} \right\|$ is bounded for all n, and by equation (18) it follows that the remainder must go to zero, that is, $\left\| \sum_{i=n}^{\infty} \beta_i y_i \right\| \to 0$ as $n \to \infty$. Therefore, $\lim_{n \to \infty} \left\| A^{-1}x - V^{(n)}x \right\| = 0$.

Notice that all these results have been established without regard to the choice of α_n. As defined in equations (10), r_n is a residual vector. The reason for this terminology will now be explained.

Suppose $r_n = \theta$ for some n. Then $V^{(n)}y_n = \alpha_n s_n$, and if $\left(V^{(n)}\right)^{-1}$ exists, $y_n = -\alpha_n\left(V^{(n)}\right)^{-1}V^{(n)}g_n = -\alpha_n g_n$ so that by definition $y_n = g^* - g_n = -\alpha_n g_n$. Since J is quadratic, $g^* = g_n + \alpha_n A s_n$. Therefore, $\alpha_n A s_n = -\alpha_n g_n$ so that $s_n = -A^{-1}g_n$. Hence, since $s_n = -V^{(n)}g_n$, $V^{(n)}g_n = A^{-1}g_n$.

Recall that the minimum of J (quadratic as in eqs. (1) and (2)) is attained by $\tilde{x} = x_n - A^{-1}g_n$. Step (2) of the basic algorithm is that if $r_n = \theta$, $\alpha_n \neq 1$; therefore, let $\alpha_n = 1$ and repeat step (1). Then the new x^* is $x^* = x_n - V^{(n)}g_n$ and it has been shown that $V^{(n)}g_n = A^{-1}g_n$. Therefore, x^* is the location of the minimum of J. This result explains the reason for step (2), and the following theorem has been proved.

Theorem 9: If $r_n = \theta$ and $\left(V^{(n)}\right)^{-1}$ exists, then by applying step (2) of the basic algorithm, α_n is set equal to 1 and the resulting x^* given by $x^* = x_n - V^{(n)}g_n$ is found to be the location of the minimum of J.

CONVERGENCE IF α_n IS CHOSEN BY A ONE-DIMENSIONAL MINIMIZATION PROCESS

There are two rather obvious ways to choose α_n at each step: (1) let $\alpha_n = 1$ for all n, and (2) let α_n be such that $J\left(x_n + \alpha_n s_n\right) \leq J\left(x_n + \lambda s_n\right)$ for all real λ. Both cases have been investigated by Davidon and Goldfarb and convergence has been established in the case of a quadratic functional on a finite-dimensional Hilbert space.

The convergence of the algorithm to the location of the minimum of a quadratic functional on an infinite-dimensional real Hilbert space when α_n is chosen for every n so that

$$J\left(x_n + \alpha_n s_n\right) \leq J\left(x_n + \lambda s_n\right) \tag{19}$$

for all real λ will now be shown. This relationship, of course, implies that $x_{n+1} = x^*$ in step (5) of the algorithm. If α_n is chosen in this manner, then by necessity

$$\frac{dJ\left(x_n + \lambda s_n\right)}{d\lambda} = 0$$

at $\lambda = \alpha_n$. Hence,

$$\alpha_n = -\frac{\left(s_n, g_n\right)}{\left(s_n, A s_n\right)} \tag{20}$$

Therefore, it must be that

$$J(x_{n+1}) = J(x_0) - \sum_{i=0}^{n} \frac{(s_i,g_i)^2}{2(s_i,As_i)}$$

Since, $\inf J > -\infty$,

$$\sum_{i=0}^{\infty} \frac{(s_i,g_i)^2}{2(s_i,As_i)} < \infty$$

which implies that by necessity

$$\lim_{i\to\infty} \frac{(s_i,g_i)^2}{(s_i,As_i)} = 0 \tag{21}$$

Since its derivation in no way depended on the method of choosing s_i, equation (21) must be true for any descent method. This well-known result and the following lemma are given by Horwitz and Sarachik (ref. 10). They used their results to prove convergence of Davidon's first method, steepest descent, and the conjugate gradient method in an infinite-dimensional real Hilbert space for the problem under consideration, namely, minimizing a quadratic functional as in equation (1).

Lemma 1: If $g_n \to \theta$ as $n \to \infty$, then x_n converges in norm to the location of the minimum $\tilde{x} = -A^{-1}b$.

Proof: $0 \leqq \left(x_n + A^{-1}b, A\left(x_n + A^{-1}b\right)\right) = \left(x_n + A^{-1}b, g_n\right) \leqq \left|\left|x_n + A^{-1}b\right|\right| \left|\left|g_n\right|\right|$. Now $\left|\left|x_n + A^{-1}b\right|\right|$ is bounded for all n, since for all n, x_n is contained in a bounded set, namely, $S_x' = \overline{\text{conv}}\{x \in H: J(x) \leqq J(x_0)\}$ (ref. 15), the closed convex hull of the indicated set. Hence $\lim_{n\to\infty} \left(x_n + A^{-1}b, A\left(x_n + A^{-1}b\right)\right) = 0$ and since A is strongly positive, $\lim_{n\to\infty} x_n + A^{-1}b = \theta$.

A general convergence theorem for this case can now be proved.

Theorem 10: If there exist positive reals α,β such that $\alpha I \leqq V^{(n)} \leqq \beta I$ for all n larger than some N and if α_n is chosen as in equation (19) then $\lim_{n\to\infty} \left|\left|x_n + A^{-1}b\right|\right| = 0$, that is, x_n converges in norm to the location of the minimum.

Proof: Since for all $u \in H$, $m||u||^2 \leqq (u,Au) \leqq M||u||^2$,

$$\frac{1}{M||u||^2} \leqq \frac{1}{(u,Au)} \leqq \frac{1}{m||u||^2}$$

and since $\alpha||u||^2 \leqq \left(u,V^{(n)}u\right) \leqq \beta||u||^2$ for all n sufficiently large,

$$\frac{1}{\beta||u||^2} \leqq \frac{1}{\left(u,V^{(n)}u\right)} \leqq \frac{1}{\alpha||u||^2}$$

Since $V^{(n)}$ is self-adjoint, $\left||V^{(n)}u\right|| \leqq \beta||u||$. Therefore,

$$\frac{(s_n,g_n)^2}{(s_n,As_n)} \geqq \frac{(s_n,g_n)^2}{M||s_n||^2} = \frac{\left(g_n,V^{(n)}g_n\right)^2}{M\left||V^{(n)}g_n\right||^2} \geqq \frac{\left(g_n,V^{(n)}g_n\right)^2}{M\beta^2||g_n||^2} \geqq \frac{\alpha^2}{M\beta^2}\frac{||g_n||^4}{||g_n||^2} = \frac{\alpha^2}{M\beta^2}||g_n||^2 \geqq 0$$

Therefore, by equation (21) $g_n \to \theta$ as $n \to \infty$ and by lemma 1 $x_n \to A^{-1}b$ in norm.

The hypothesis of theorem 10 is more restrictive than necessary, only $\alpha||g_n||^2 \leqq \left(g_n,V^{(n)}g_n\right) \leqq \beta||g_n||^2$ is required. A result similar to that of theorem 10 but more general can be found in reference 7. Theorem 5 can now be combined with theorem 10 to state a convergence theorem for the rank-one iterates.

Corollary 4: If $V^{(0)} \leqq A^{-1}$ or $V^{(0)} \geqq A^{-1}$ and α_n is chosen as in equation (19), then $J(x_n)$ converges to the minimum of $J(x)$, and moreover x_n converges in norm to the location of the minimum.

Proof: If $V^{(0)} \leqq A^{-1}$, then by theorems 5 and 6, $V^{(0)} \leqq V^{(n)} \leqq A^{-1}$ for all n. Hence, $m_0 I \leqq V^{(n)} \leqq \frac{1}{m} I$ for all n.

CONVERGENCE WITH A MORE GENERAL CHOICE OF α_n

In this section, as in the previous one, the convergence of rank-one iterates to the location of the minimum of a quadratic functional as in equations (1) and (2) is investigated. However, herein the step size α_n need not be chosen by a one-dimensional minimization. Conditions on the initial estimate and the α_n scalars under which the iterates converge to $\tilde{x} = -A^{-1}b$, the location of the minimum, are determined.

Let $\{\alpha_i\}$ denote the sequence of step sizes used by the rank-one algorithm. (From eq. (8), $x^* = x_i + \alpha_i s_i$ for each i.) For each $n \geqq 1$, denote by $N(n)$ that set of integers less than n for which $J\left(x_i - \alpha_i V^{(i)}g_i\right) < J(x_i)$ and the integer $n - 1$; that is, $N(n)$ contains the indices of those step sizes which caused a decrease in the function being minimized and the integer $n - 1$. If $y_i = g_{i+1} - g_i$, then

$$g_n = g_0 + \sum_{i \in N(n)} y_i \tag{22}$$

Hence,

$$V^{(n)}g_n = V^{(n)}g_0 + \sum_{i \in N(n)} V^{(n)}y_i = V^{(n)}g_0 + \sum_{i \in N(n)} (x_{i+1} - x_i) \tag{23}$$

From corollary 2, $V^{(n)}y_i = \alpha_i s_i = x_{i+1} - x_i$ for all $i < n$, and hence for all $i \in N(n)$. If $x_{i+1} - x_i$ is denoted by σ_i, then by step (5) of the algorithm and the definition of $N(n)$

$$x_n = x_0 + \sum_{i \in N(n)} \sigma_i \tag{24}$$

From equations (8), (23), and (24),

$$x^* = x_n - \alpha_n V^{(n)}g_n = x_0 + \sum_{i \in N(n)} \sigma_i - \alpha_n \left(V^{(n)}g_0 + \sum_{i \in N(n)} \sigma_i \right)$$

Hence,

$$x^* = x_0 - \alpha_n V^{(n)}g_0 + (1 - \alpha_n) \sum_{i \in N(n)} \sigma_i \tag{25}$$

Then by using equation (25)

$$\left\| x^* + A^{-1}b \right\| \leq \left\| \left(A^{-1} - \alpha_n V^{(n)} \right) g_0 \right\| + \left\| (1 - \alpha_n) \sum_{i \in N(n)} \sigma_i \right\| \tag{26}$$

In order to establish convergence, it must be shown that $\left\| x^* + A^{-1}b \right\|$ can be made small as $n \to \infty$. Let $S'_{x_0} = \operatorname{conv}\{x \in H : J(x) \leq J(x_0)\}$. Since it is known that S'_{x_0} is bounded (ref. 15), the following lemma can be proved.

Lemma 2: If $(\alpha_n - 1)n \to 0$ as $n \to \infty$ and there exist α, $\beta > 0$ such that $\alpha I \leq V^{(n)} \leq \beta I$ and $\tau_n \neq 0$ for all n, then $\left\| (1 - \alpha_n) \sum_{i \in N(n)} \sigma_i \right\| \to 0$ as $n \to \infty$.

Proof: By definition $||\sigma_i|| = ||\alpha_i s_i|| = ||-\alpha_i V^{(i)} g_i||$. Therefore

$$||\sigma_i|| = |\alpha_i| \, ||V^{(i)}(Ax_i + b)|| \leq |\alpha_i| \left\{ ||V^{(i)}|| \cdot ||A|| \, ||x_i|| + ||V^{(i)}|| \cdot ||b|| \right\} \qquad (27)$$

Since $x_i \in S'_{x_0}$ is a bounded set, $||x_i||$ is bounded and since $\alpha_i \to 1$ as $i \to \infty$, α_i is bounded. By hypothesis $||V^{(i)}|| \leq \beta$ and $||A|| \leq M$; thus, everything on the right-hand side of equation (27) is independent of i and $||\sigma_i|| \leq L$ for some $L \geqq 0$ and all i. Hence,

$$\left|\left|(1 - \alpha_n) \sum_{i \in N(n)} \sigma_i \right|\right| \leq (1 - \alpha_n) n \cdot L \to 0$$

since by hypothesis $(\alpha_n - 1)n \to 0$ as $n \to \infty$.

Lemma 3: Suppose either g_0 is an element of the smallest subspace containing the y_i vectors denoted by $S(y_i)$ or the y_i vectors form a Schauder basis for H. If the $V^{(n)}$ operators are uniformly bounded, $a_n \to 1$ as $n \to \infty$, and $\tau_n \neq 0$ for all n, then $\left|\left|\left(A^{-1} - \alpha_n V^{(n)}\right) g_0\right|\right| \to 0$ as $n \to \infty$.

Proof: Write α_n as $1 - \epsilon_n$; then since $\alpha_n \to 1$ as $n \to \infty$, $\epsilon_n \to 0$ as $n \to \infty$. Therefore,

$$\left|\left|\left(A^{-1} - \alpha_n V^{(n)}\right) g_0\right|\right| = \left|\left|A^{-1} g_0 - V^{(n)} g_0 + \epsilon_n V^{(n)} g_0\right|\right| \leq \left|\left|A^{-1} g_0 - V^{(n)} g_0\right|\right| + |\epsilon_n| \left|\left|V^{(n)} g_0\right|\right|$$

Hence because of the hypothesis on g_0, $\lim_{n \to \infty} \left|\left|A^{-1} g_0 - V^{(n)} g_0\right|\right| = 0$ from theorem 7 or 8. Also since the $V^{(n)}$ operators are uniformly bounded, $\left|\left|V^{(n)} g_0\right|\right|$ must be bounded for all n. Therefore $\lim_{n \to \infty} |\epsilon_n| \left|\left|V^{(n)} g_0\right|\right| = 0$. Hence the lemma is proved.

The following statement can now be made.

Theorem 11: Suppose $g_0 \in S(y_i)$ or the y_i vectors form a Schauder basis for H. If $\tau_n \neq 0$ for all n, the $V^{(n)}$ are uniformly bounded, and if $(\alpha_n - 1)n \to 0$ as $n \to \infty$, then $||x^* + A^{-1} b|| \to 0$ as $n \to \infty$.

Proof: By equation (26)

$$\left|\left|x^* + A^{-1} b\right|\right| \leq \left|\left|\left(A^{-1} - \alpha_n V^{(n)}\right) g_0\right|\right| + \left|\left|(1 - \alpha_n) \sum_{i \in N(n)} \sigma_i\right|\right|$$

and by lemma 3 the first term goes to zero. By lemma 2 the second term goes to zero.

Hence, conditions under which two variations of the basic algorithm converge to the location of the minimum of a quadratic functional have been established. These conditions are given in theorems 10 and 11. In both of these theorems, the primary interest is in the convergence question for an infinite-dimensional Hilbert space. In a finite-dimensional space of dimension n, it can be seen that for almost any collection of non-zero α_n scalars, the algorithm converges to the location of the minimum in a finite number of steps. The conditions on the α_n scalars and the proof are given in the following theorem from Broyden (ref. 1).

Theorem 12: If $\tau_j \neq 0$ and $\alpha_j \neq 0$ for all $j = 0, 1, \ldots$ and if $\left(V^{(j)}\right)^{-1}$ exists for all j, then after at most $q + 1$ steps $x^* = -A^{-1}b$ where $q = \dim(H)$.

APPLICATION TO OPTIMAL CONTROL THEORY

Fixed-time problems will be considered since by a simple transformation (ref. 17), the free-time problem can be transformed into a fixed-time problem. Moreover, Horwitz and Sarachik (ref. 21) have given several other schemes for solving the free-time problem by using fixed-time techniques, and these schemes are applicable when the Davidon-Broyden algorithms are used.

A Quadratic Payoff With Linear Constraining Differential Equations

From the class $\mathcal{L}_r^2(t_0,t_1)$, the problem is to find a function $\tilde{u}(t)$ which minimizes

$$J(u) = \frac{1}{2}\int_{t_0}^{t_1} \left\{x(t)^T P(t)\, x(t) + u(t)^T R(t)\, u(t)\right\} dt \tag{28}$$

subject to the constraints

$$\dot{x}(t) = G(t)\, x(t) + B(t)\, u(t) \tag{29}$$

and $x(t_0) = x_0$ where x_0, t_0, and t_1 are fixed. For this problem,

x — n vector

u — r vector

G(t) — $n \times n$ matrix with components in $L^1(t_0,t_1)$

B(t) — $n \times r$ matrix with components in $L^1(t_0,t_1)$ and bounded

$P(t)$ $n \times n$ symmetric, positive semidefinite matrix the components of which are piecewise continuous on (t_0,t_1)

$R(t)$ $r \times r$ symmetric, uniformly positive definite matrix the components of which are piecewise continuous on (t_0,t_1)

Horwitz and Sarachik (ref. 10) have shown that this problem can be considered as that of finding the location of the minimum of a quadratic functional on $\mathcal{L}_r^2(t_0,t_1)$. This function is exactly the type of function for which the conditions given in theorems 10 and 11 guarantee the convergence of the various modifications of the basic algorithm.

General Optimal Control Problems and Gradient of Payoff

In this section a class of problems generally referred to as optimal control problems (ref. 22) or in the calculus of variations as Lagrange problems (ref. 17) is described. Also presented formally are the mechanics of applying the algorithms discussed to compute solutions to these problems, even though these functions generally are not quadratic.

Given a system of n differential equations such as

$$\dot{x}(t) = f(x,u,t) \tag{30}$$

with $x(t_0) = x_0$ and $u \in R^r$, find that function $u = \tilde{u}(t)$ which minimizes the value of $\int_{t_0}^{t_1} L(x(t),u(t),t)\,dt$. Assume that $f(x,u,t)$ and $L(x,u,t)$ have continuous partial derivatives of at least second order in x and u and are piecewise continuous in t. Also, assume that there are no constraints on u or x, other than that x must satisfy equation (30).

Moreover, it will be assumed that L and f are such that corresponding to every $u = u(t) \in \mathcal{L}_r^2(t_0,t_1)$, a real Hilbert space, there exists a solution, $x = x(t)$, of equation (30) and that for this x and u the integral $\int_{t_0}^{t_1} L(x(t),u(t),t)\,dt$, exists. Hence, the functional $J:H \to R$ can be defined by

$$J(u) = \int_{t_0}^{t_1} L(x(t),u(t),t)\,dt$$

where $x(t)$ is a solution of equation (30) corresponding to u.

Therefore, the problem is that of locating the minimum of a functional J on a real Hilbert space H. In order to apply the algorithms discussed earlier, the gradient of J must be computed. The gradient of J is that part of $J(u + \delta_u) - J(u)$ which is linear in δ_u.

It can be shown (ref. 23) that if the Hamiltonian is defined as

$$H(x,\lambda,t,u) = L(x,u,t) + \lambda^T f(x,u,t) \tag{31}$$

then the gradient of J at u is given by

$$g(u) = \nabla_u H = \left[L_u(x(t),u(t),t) + \lambda^T f_u(x(t),u(t),t)\right]^T \tag{32}$$

where

$$\dot{x} = f(x(t),u(t),t) = \frac{\partial H}{\partial \lambda} \qquad \left(x(t_0) = x_0\right)$$

$$\dot{\lambda} = -\frac{\partial H}{\partial x}(x(t),\lambda(t),t,u(t)) \qquad \left(\lambda(t_1) = 0\right)$$

The steps necessary to compute the gradient of J at $u = u_0(t)$ are (1) numerically integrate $\dot{x} = f(x,u_0,t)$ with $x(t_0) = x_0$ forward to $t = t_1$, and (2) at $t = t_1$ integrate (again numerically)

$$\dot{\lambda} = -f_x^T(x,u_0,t)\lambda - L_x^T(x,u_0,t)$$

with $\lambda(t_1) = 0$ backward to $t = t_0$. Therefore, the gradient as given in equation (32) can be computed by using the control $u = u_0(t)$ and the values of $x(t)$ and $\lambda(t)$ previously computed. In practice, u_0 is usually expressed in tabular form. If the gradient is computed according to equation (32), then $B^{(n)}$ and r_n can be computed as in equations (13) and (10). Hence, the algorithms outlined can be used to generate a sequence of controls which hopefully converge to the optimal. The practical significance of theorem 11 is to demonstrate that the rank-one algorithms are not as sensitive to the accuracy of the one-dimensional minimization as are the other widely used gradient algorithms, for example, conjugate gradient and Davidon-Fletcher-Powell (DFP) methods. For minimizing many functions (not necessarily quadratic) by these two algorithms, a large part of the computing time is required to solve the one-dimensional minimization problem repeatedly.

EXAMPLE PROBLEM AND RESULTS OF ANALYSIS

In order to exhibit the convergence characteristics of the rank-one algorithms, a sample optimal control problem which others have used to display convergence character-

istics of other algorithms (refs. 8, 9, and 11) has been chosen. The problem is to find the function $u = u(t)$ which minimizes

$$J = \int_0^5 \left(x_1^2 + x_2^2 + u^2\right) dt \tag{33}$$

subject to constraining differential equations described by the Van der Pol equation (ref. 24) with $\epsilon = 1$, that is,

$$\left.\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -x_1 + \left(1 - x_1^2\right)x_2 + u \end{aligned}\right\} \tag{34}$$

with initial conditions

$$x_1(0) = 3.0$$

$$x_2(0) = 0.0$$

By equation (32) the gradient g of J at u is given by

$$g(t) = 2u(t) + \lambda_2(t) \tag{35}$$

where

$$\left.\begin{aligned} \dot{\lambda}_1 &= \left(1 + 2x_1x_2\right)\lambda_2 - 2x_1 \\ \dot{\lambda}_2 &= -\lambda_1 - \left(1 - x_1^2\right)\lambda_2 - 2x_2 \end{aligned}\right\} \tag{36}$$

with

$$\lambda_1(5) = 0$$

$$\lambda_2(5) = 0$$

In order to compute the gradient $g(t)$ of J at some t, equations 34 are integrated forward to $t = 5.0$ by using $u = u_0(t)$. Next, equations (36) are integrated from $t = 5.0$ back to $t = 0.0$. Then by using $u = u_0(t)$ and the computed value of $\lambda_2(t)$, $g(t)$ given by equation (35) can be computed.

Figures 1 and 2 depict the progress toward the minimum of J by using the algorithm outlined with four different methods of choosing α_n. These four methods of choosing α_n are:

(1) $\alpha_n = 1 - (n^3 + 2)^{-1/2}$ for all n.

(2) $\alpha_n = 1$ for all n.

(3) $\alpha_n = \min\left\{1.0, \left[\tilde{J}_m - J(u_n)\right] / (s_n, g_n)\right\}$ where $\tilde{J}_m$ is the estimated minimum value of J, s_n is defined by equation (9), and g_n is the gradient of J at $u = u_n(t)$.

(4) α_n is the minimum with respect to α of $J(u_n + \alpha s_n)$ as computed by Davidon's one-dimensional cubic minimization method (ref. 2).

Methods (1) and (2) of choosing α_n satisfy the condition given in theorem 11 (for these methods $(1 - \alpha_n)n \to 0$ as $n \to \infty$). The form of α_n for method (3) follows by considering $\tilde{J}_m = J(u_n) + \alpha_n(s_n, g_n)$ + higher order terms, dropping the higher order terms, and solving for α_n, where $\tilde{J}_m$ is some nominal estimate of the minimum of J along the line $u_n + \alpha s_n$.

Notice that methods (1), (2), and (3) of choosing α_n involve no extra functional and gradient evaluations; that is, for each iteration equations (34) and (36) must be integrated only once. The fourth method of choosing α_n, although the one-dimensional minimum is computed more accurately than by method (3), involves at least one more functional evaluation per iteration. Hence, with the fourth method of choosing α_n, at least two functional and gradient evaluations per iteration are required.

In figure 2, $J(u_n)$ is plotted against n (that is, the iteration number) for the four different methods of choosing α_n. Figure 2 shows that the fastest convergence in terms of iterations is achieved by the algorithm with α_n chosen by method (4). Also, figure 2 shows that after 12 iterations, all the methods have converged. Moreover, after eight iterations for all methods of choosing α_n, the change in the value of J is too small to show up in the graph.

In figure 3, J is plotted against the number of functional evaluations. Notice that in figure 3, methods (3) and (1) converge faster with respect to function evaluations than method (4). Note also that after at most eight functional evaluations, the change in J is too small to be noticed in the graph.

Figure 4 shows the rates of convergence to the minimum for the example problem for the three first-order methods (steepest descent, conjugate gradient, and DFP). These results were reported by Tokumaru, et al. (ref. 11). Note that the DFP algorithm shows the fastest rate of convergence.

By using the same initial estimate of u that was used for the results shown in figure 2, the DFP method was applied to the example problem. The results for the DFP

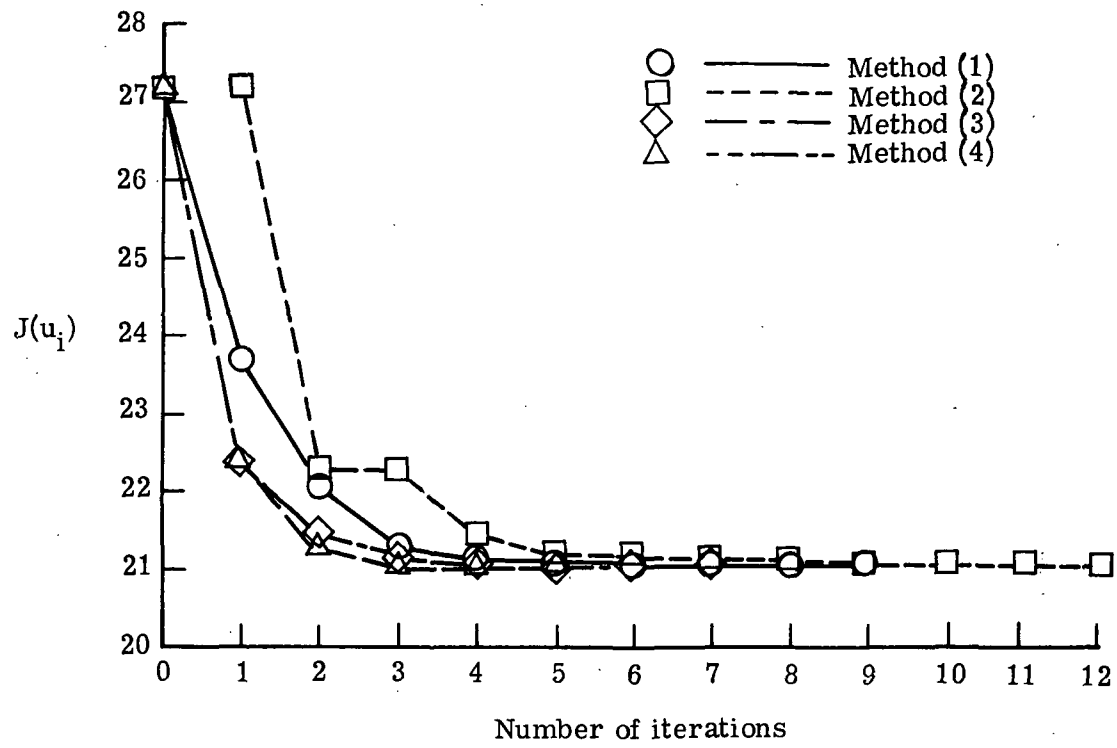

Figure 2.- Variation of $J(u_i)$ with i for the four methods of choosing α_i.

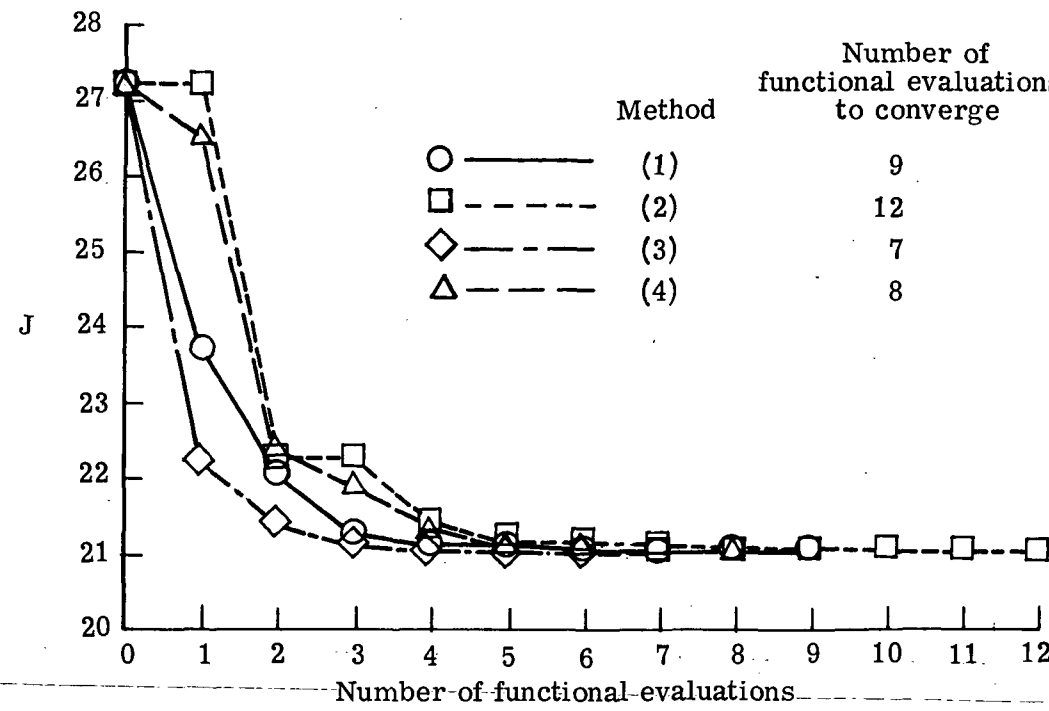

Figure 3.- Rank-one performance for various step size criteria.

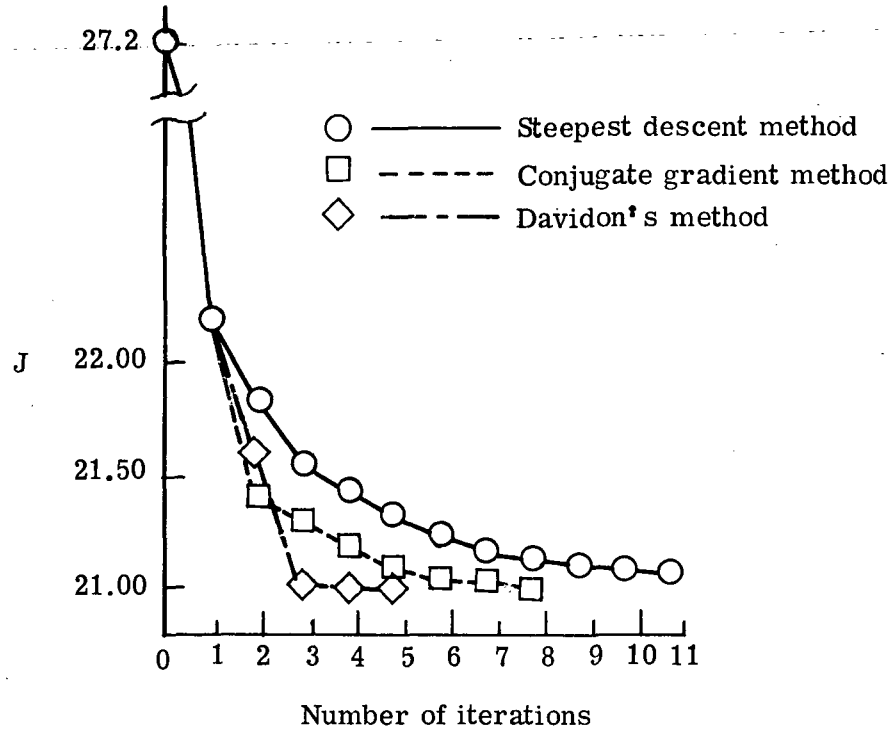

Figure 4.- Comparison of previous methods. From reference 11.

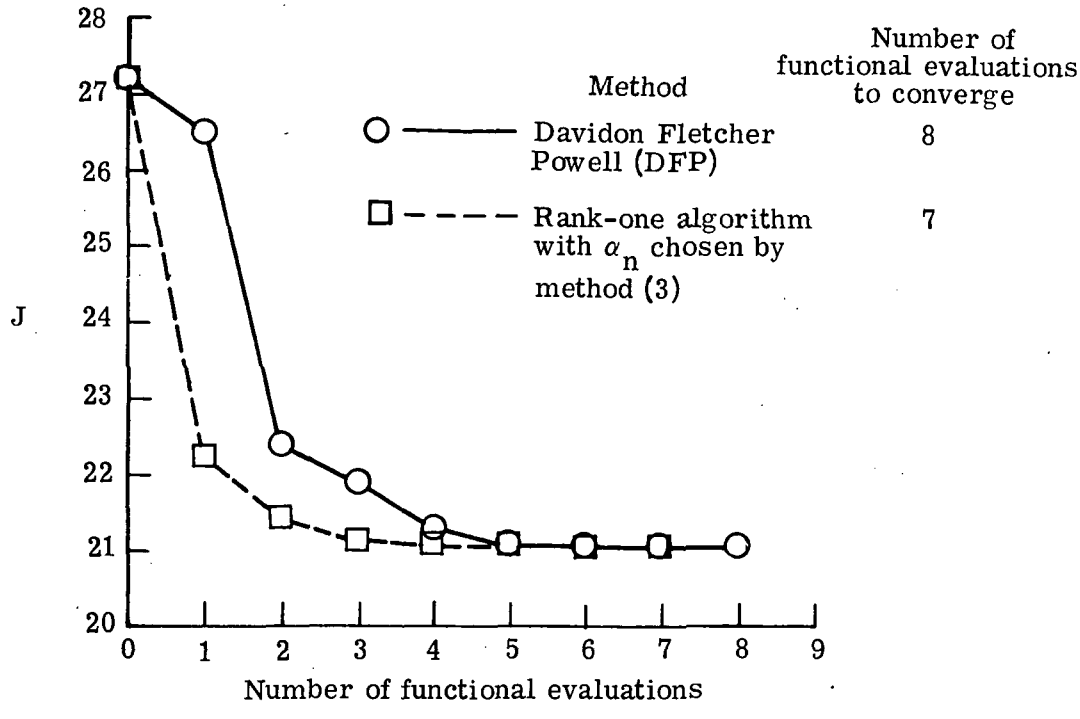

Figure 5.- Rank-one performance compared with Davidon-Fletcher-Powell (DFP) method.

method were identical to those of the rank-one algorithm with α_n chosen by method (4). The reduction in the payoff and the iterates for the two methods were the same.

In figure 5, the values of $J(u_n)$ have been plotted against the number of function evaluations for the DFP method and the algorithm when α_n is chosen by method (3). Notice that in terms of function evaluations, the rank-one method for this choice of α_n converges faster than the DFP algorithm. The linear minimizations for the DFP algorithm were carried out by method (4). This method was chosen because high accuracy in the linear minimization is necessary for the DFP method.

The rank-one algorithms outlined herein have several attractive properties. Theorem 7 gives conditions under which $V^{(n)} \to A^{-1}$ pointwise where $V^{(n)}$ is given by equation (12) and A is given by equation (2). This property can be used to accelerate the convergence when many solutions corresponding to different initial conditions are desired. This property is not available with the method of conjugate gradients. Moreover, these rank-one algorithms require one-half the storage necessary for the DFP method. Also, rank-one methods require the computation of only one operator per iteration whereas the computation of two operators are required per DFP iteration.

The results of the example problem show that the algorithm can be applied with success when α_n is chosen in a variety of ways. It appears that method (3) of choosing α_n is best when the functional to be evaluated is very complex, its computation is time consuming, and storage considerations are not as important. If storage considerations are pressing and the computation of the functional is not as time consuming, then method (4) would seem to be the best choice for α_n.

CONCLUDING REMARKS

The extension to a real Hilbert space of the rank-one minimization algorithms has been presented. Conditions insuring the convergence to the location of the minimum of a quadratic functional are given for various techniques of choosing the step size. Also the application of these algorithms for the direct computation of optimal controls has been outlined. An example optimal control problem has been solved by using several rank-one algorithms. The example problem shows that rank-one algorithms give promise of a fast rate of convergence, comparable to that for the Davidon-Fletcher-Powell (DFP) method. Moreover, the rank-one techniques use less storage and involve fewer computations than the DFP method. Also, the example problem illustrates that the one-dimensional search required by DFP is not critical for convergence when a rank-one algorithm is used. In

fact, for the example problem, convergence to the location of the minimum was achieved by using four different methods for choosing the step size. Two of the choices for step size used in the example were independent of the function values and yet convergence was achieved.

Langley Research Center,
National Aeronautics and Space Administration,
Hampton, Va., October 26, 1972.

REFERENCES

1. Broyden, C. G.: A Class of Methods for Solving Nonlinear Simultaneous Equations. Math. Comput., vol. 19, no. 92, Oct. 1965, pp. 577-593.

2. Davidon, William C.: Variable Metric Method for Minimization. ANL 5990 Rev. (Contract W-31-109-eng-38), Argonne Nat. Lab., Nov. 1959.

3. Fletcher, R.; and Powell, M. J. D.: A Rapidly Convergent Descent Method for Minimization. Computer J., vol. 6, no. 2, July 1963, pp. 163-168.

4. Fletcher, R.; and Reeves, C. M.: Function Minimization by Conjugate Gradients. Computer J., vol. 7, no. 2, July 1964, pp. 149-154.

5. Hestenes, Magnus R.; and Stiefel, Edward: Methods of Conjugate Gradients for Solving Linear Systems. Res. Paper 2379, J. Res. Nat. Bur. Stand., vol. 49, no. 6, Dec. 1952, pp. 409-436.

6. Daniel, James W.: The Conjugate Gradient Method for Linear and Nonlinear Operator Equations. SIAM J. Numer. Anal., vol. 4, no. 1, Mar. 1967, pp. 10-26.

7. Daniel, James W.: The Approximate Minimization of Functionals. Prentice-Hall, Inc., c.1971.

8. Lasdon, L. S.; Mitter, S. K.; and Waren, A. D.: The Conjugate Gradient Method for Optimal Control Problems. IEEE Trans. Automat. Contr., vol. AC-12, no. 2, Apr. 1967, pp. 132-138.

9. Sinnott, J. F., Jr.; and Luenberger, D. G.: Solution of Optimal Control Problems by the Method of Conjugate Gradients. 1967 Joint Automatic Control Conference, June 1967, pp. 566-573.

10. Horwitz, Lawrence B.; and Sarachik, Philip E.: Davidon's Method in Hilbert Space. SIAM J. Appl. Math., vol. 16, no. 4, July 1968, pp. 676-695.

11. Tokumaru, H.; Adachi, N.; and Goto, K.: Davidon's Method for Minimization Problems in Hilbert Space With An Application to Control Problems. SIAM J. Control, vol. 8, no. 2, May 1970, pp. 163-178.

12. Davidon, William C.: Variance Algorithms for Minimization. Computer J., vol. 10, no. 4, Feb. 1968, pp. 406-410.

13. Davidon, William C.: Variance Algorithms for Minimization. Optimization, R. Fletcher, ed., Academic Press, 1969, pp. 13-20.

14. Goldfarb, D.: Sufficient Conditions for the Convergence of a Variable Metric Algorithm. Optimization, R. Fletcher, ed., Academic Press, 1969, pp. 273-283.

15. Luenberger, David G.: Optimization by Vector Space Methods. John Wiley & Sons, Inc., c.1969.

16. Lusternik, L. A.; and Sobelev, V. J.: Elements of Functional Analysis. Hindustan Pub. Corp. (Delhi, India), c.1961. (Available from Gordon & Breach Publ., Inc., New York.)

17. Sagan, Hans: Introduction to the Calculus of Variations. McGraw-Hill Book Co., Inc., c.1969.

18. Akhiezer, N. I.; and Glazman, I. M. (Merlynd Nestell, trans.): Theory of Linear Operators in Hilbert Space. Vol. I. Frederick Ungar Pub. Co., c.1961.

19. Berberian, Sterling K.: Introduction to Hilbert Space. Oxford Univ. Press, 1961.

20. Goffman, Casper; and Pedrick, George: First Course in Functional Analysis. Prentice-Hall, Inc., c.1965.

21. Horwitz, L. B.; and Sarachik, P. E.: A Survey of Two Recent Iterative Techniques for Computing Optimal Control Signals. Tenth Joint Automatic Control Conference of the American Automatic Control Council, Aug. 1969, pp. 50-51.

22. Pontryagin, L. S.; Boltyanskii, V. G.; Gamkrelidze, R. V.; and Mishchenko, E. F.: The Mathematical Theory of Optimal Processes. Interscience Publ., Inc., c.1962.

23. Bryson, Arthur E., Jr.; and Ho, Yu-Chi: Applied Optimal Control. Blaisdell Pub. Co., c.1969.

24. Struble, Raimond A.: Nonlinear Differential Equations. McGraw-Hill Book Co., Inc., 1962.

NASA-Langley, 1972 — 19 L-8576

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE $300

SPECIAL FOURTH-CLASS RATE
BOOK

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION
451

U.S.MAIL

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons. Also includes conference proceedings with either limited or unlimited distribution.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include final reports of major projects, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION OFFICE

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546