A GLOBALLY CONVERGING SECANT METHOD, WITH APPLICATIONS TO BOUNDARY VALUE PROBLEMS

by

E. Polak

Memorandum No. ERL-M374

18 September 1972

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A Globally Converging Secant Method, With Applications

to Boundary Value Problems

E. Polak

Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California, Berkeley, California 94720

1. Introduction.

In this paper we present a new algorithm, in the secant methods family, for solving equations of the form $g(z) = 0$, where $g : \mathbb{R}^n \to \mathbb{R}^n$ is twice continuously differentiable and its Jacobian matrix $\frac{\partial}{\partial z} g(z)$ is invertible. Under mild assumptions, our algorithm will converge to a solution $\hat{z}$, irrespective of whether the initial guess z_0 is a good approximation to $\hat{z}$ or not. After a few iterations, our algorithm requires only two function evaluations per iteration and the iterates z_i which it constructs satisfy $\|z_i - \hat{z}\| \leq K\theta^{\tau_n^i}$ for all $i \geq i_0$, $K > 0$, $\theta \in (0,1)$, with τ_n being the unique positive root of $t^{n+1} - t^n - 1 = 0$. By (9.2.8) in [6], $1 < \tau_n < 2$ and $\tau_n \downarrow 1$ as $n \to \infty$. Hence the efficiency of our algorithm, η, defined as the ratio of rate (τ_n) to the number of function evaluations per iteration, is seen to be $\tau_n/2n$, and hence $\frac{1}{2n} < \eta < \frac{1}{n}$. We note that the efficiency of the Newton method is $2/(n^2+n)$ and hence that our method is considerably more efficient than Newton's method, especially when n is large. The superiority of our method is in fact even greater, because of the much smaller cost involved

---

Research sponsored by the NASA NGL-05-003-016/JSEP F44620-71-C-0087/ NSF GK-10656X2.

in matrix inversions.

The algorithm in this paper utilizes four ideas, represented by the sequential secant methods of Wolfe [9] and Barnes [4], the results for secant methods based on consistent approximations in Section 11.2 of Ortega and Rheinboldt [6], the method of local variations of Banitchouk, Petrov and Chernousko [3], and the convergence theory described in Section 1.3 of Polak [7]. The result is a globally convergent algorithm with the ideal rate of convergence of a sequential secant method. We note that it is superior both to Newton's method, because it is more efficient, and to the above mentioned secant methods because it is stable and globally convergent and they are not. Among the more interesting applications we foresee for our method is in the solution of moderately well behaved, two point boundary value problems. We have used it in this context and have found it to behave very well.

2. The Secant Method.

Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be a twice continuously differentiable function, whose Jacobian matrix will be denoted by $G(z)$, i.e., $G(z) = \frac{\partial}{\partial z} g(z)$. We shall need the following

1. Assumptions:

(i) There is a $z_0 \in \mathbb{R}^n$ such that the set $C(z_0) = \{z \mid \|g(z)\| \leq \|g(z_0)\|\}$ is compact.

(ii) The set $C(z_0)$, above, contains at least one point $\hat{z}$ such that $g(\hat{z}) = 0$, and the number of points $\hat{z} \in C(z_0)$, such that $g(\hat{z}) = 0$, is finite.

(iii) Let $S = \{\hat{z} \in C(z_0) \mid g(\hat{z}) = 0\}$, then there exist an $L > 0$ and a $\rho > 0$ such that $\|G(z') - G(z'')\| \leq L \|z' - z''\|$ $\forall z', z'' \in B(\hat{z},\rho)$, $\forall \hat{z} \in S$, where

$B(\hat{z},\rho) = \{z \mid \|z - \hat{z}\| \leq \rho\}$.†

(iv) $G(z)^{-1}$ exists and is continuous for all z in an open set containing $C(z_0)$.

Under the assumptions stated, we will show that the algorithm below will construct a sequence $\{z_i\}$ which converges superlinearly to a point $\hat{z}$ satisfying $g(\hat{z}) = 0$.

3. Algorithm.

Data: $\delta > 0$, $\alpha \in (0, \frac{1}{6})$, $\beta \in (0,1)$, $b > 0$ (large), $\ell \geq 1$, $z_0 \in \mathbb{R}^n$, $H \in L(n)$‡; e_j = jth column of nxn unit matrix, $j = 1,2,\ldots,n$.

Step 0: For $m = 1,2,\ldots,n$, set $d_m = e_m$, $d_{n+m} = - e_m$; set $i = 0$, $j = 0$, $s = 0$, $\nu = \infty$, $\bar{H} = H$.

Step 1: Compute $g(z_i)$.

Step 2: If $j < 2n$, set $j = j + 1$ and go to Step 3; else, set $j = 1$ and go to step 3.

Step 3: Set $\varepsilon_i = \min\{\delta,\nu\}$.

Step 4: Compute $g(z_i+\varepsilon_i d_j)$.

Step 5: If $j \leq n$, replace $\bar{h}_j$, the jth column of $\bar{H}$, by

4. $$\Delta_i = \frac{1}{\varepsilon_i} [g(z_i+\varepsilon_i d_j) - g(z_i)],$$

else, replace $\bar{h}_j$, the (j-n) column of $\bar{H}$, by $-\Delta_i$, to obtain a new matrix (again denoted by $\bar{H}$),

5. $$\bar{H} = (\bar{h}_1\ \bar{h}_2,\ \ldots,\ \bar{h}_{j-1},\ (\pm)\Delta_i,\ \bar{h}_{j+1},\ \ldots,\ \bar{h}_n).$$

Step 6: If $\|g(z_i+\varepsilon_i d_j)\|^2 < \|g(z_i)\|^2$, set $s = 0$, $\omega = z_i + \varepsilon_i d_j$,

---

†This assumption is obviously redundant since it follows from the continuous differentiability of $G(\cdot)$ and assumption (iii). We state it simply for the sake of convenience later.

‡$L(n)$ is the space of real nxn matrices.

and go to Step 7; else, set $s = s + 1$, $\omega = z_i$, and go to Step 7.

Step 7 : If $\bar{H}^{-1}$ exists† and $\|\bar{H}^{-1}\| \le b$, set $H_i = \bar{H}$, compute

6. $$v = H_i^{-1} g(z_i)$$

and go to Step 8; else, go to Step 12.

Step 8 : Set $k = 0$.

Step 9: Compute $g(z_i - \beta^k v)$.

Step 10: If

7. $$\|g(z_i - \beta^k v)\|^2 \le (1 - 2\beta^\ell \alpha) \|g(z_i)\|^2$$

set $z_{i+1} = z_i - \beta^k v$,

$\nu = \beta^k \|v\|$, $s = 0$, $i = i+1$, and go to Step 2; else, go to Step 11.

Step 11: If $k < \ell$, set $k = k + 1$, and go to step 9, else, go to Step 12.

Step 12: If $s < 2n$, go to Step 13; else, set $s = 0$, $\delta = \delta/2$ and go to Step 13.

Step 13; If $\omega = z_i$, go to step 2; else, set $H_i = \bar{H}$, $z_{i+1} = \omega$, $i = i+1$ and go to step 2.

In constructing the above algorithm, we thought of our problem as being $\min\{\frac{1}{2}\|g(z)\|^2 | z \in \mathbb{R}^n\}$, rather than as that of finding a zero of $g(\cdot)$. Our algorithm uses the method of local variations (see [7],p.43) to construct approximations H_i to the Jacobian $G(z_i)$ and to ensure that the iterates z_i proceed towards a zero of $g(\cdot)$. After a small number

---

†Note that since $\bar{H}_{new}$ differs from $\bar{H}_{old}$ by the jth column only, whenever $\bar{H}_{new}^{-1}$ and $\bar{H}_{old}^{-1}$ exist, $\bar{H}_{new}^{-1}$ can be computed according to the formula $\bar{H}_{new}^{-1} = H_{old}^{-1} + \frac{1}{c_j \Delta_i} (e_j - \bar{H}_{old}^{-1} \Delta_i) c_j$, where c_j is the jth row of $\bar{H}_{old}^{-1}$.

of iterations, the approximations H_i become sufficiently good for the algorithm to continue in secant mode. The test (7) is a minor modification of the Armijo [2] step size rule for gradient methods and is used to ensure convergence once the algorithm enters the secant mode of operation. We shall now make the preceding statements precise.

8. Proposition: Suppose Assumptions (1(i), 1(ii) and 1(iv)) hold and that the algorithm (3) has constructed the points $z_1, z_2, \ldots, z_i$. If $g(z_i) \neq 0$, then, after at most a finite number of halvings of δ in Step 12, the algorithm will construct a point z_{i+1}, with $\|g(z_{i+1})\| < \|g(z_i)\|$.

Proof: If H_i^{-1} exists and for $k \leq \ell$ (7) can be satisfied, then the proposition follows directly. If either H_i^{-1} does not exist and/or for $k \leq \ell$ (7) cannot be satisfied, then the algorithm becomes the method of local variations, and the proposition follows from the fact that this method jams up only at points z_i satisfying $G(z_i)^T g(z_i) = 0$. (see [7],p.43).

9. Proposition: Suppose that the algorithm (3) has constructed an infinite sequence $\{z_i\}_{i=0}^{\infty}$. If there exists an infinite subset of the integers, K, such that (7) is satisfied for all $i \in K$, then $g(z_i) \to 0$ as $i \to \infty$.

Proof: The sequence $\{\|g(z_i)\|^2\}_{i=0}^{\infty}$ is, by construction, strictly monotomically decreasing and bounded from below. Hence there exists a $\gamma^* \geq 0$ such that $\|g(z_i)\|^2 \to \gamma^*$ as $i \to \infty$. Suppose that $\gamma^* > 0$, then, for any $\tau > 0$, there exists an $N \geq 0$ such that for all $i \in K$, $i \geq N$,

10. $$\|g(z_{i+1})\|^2 \leq (1-2\beta^\ell\alpha)\ \|g(z_i)\|^2 \leq (1-2\beta^\ell\alpha)(\gamma^*+\tau).$$

Since $\tau > 0$ is arbitrary, it is clear that (10) contradicts the convergence of $\|g(z_i)\|^2$ to $\gamma^* > 0$. Hence we must have $\gamma^* = 0$.

11. <u>Proposition</u>: Suppose that Assumptions (1(i), 1(ii) and 1(iv)) hold. Suppose that algorithm (3) has constructed an infinite sequence $\{z_i\}_{i=0}^{\infty}$ and that there exists an integer N such that for all $i \geq N$, either H_i^{-1} does not exist or the test (7) fails for $k = 0,1,2,\ldots \ell$. Then $\|g(z_i)\|^2 \to 0$ as $i \to \infty$.

<u>Proof</u>: For $i \geq N$, the algorithm becomes the method for local variations. Since $C(z_0)$ is compact, by the properties of the method of local variations ([7], p. 43), $\{z_i\}_{i=0}^{\infty}$ must have at least one accumulation point $\hat{z}$ which satisfies $\frac{\partial}{\partial z}\frac{1}{2}\|g(z)\|^2 = G(\hat{z})^T g(\hat{z}) = 0$. Since $G(\hat{z})^{-1}$ exists, by assumption, $g(\hat{z}) = 0$. Hence, since $\{\|g(z_i)\|^2\}_{i=0}^{\infty}$ is a monotonically decreasing, bounded sequence, we must have $\|g(z_i)\|^2 \to 0$ as $i \to \infty$.

12. <u>Theorem</u>: Suppose that assumptions (1(i), 1(ii) and 1(iv)) hold and that algorithm (3) has constructed an infinite sequence $\{z_i\}_{i=0}^{\infty}$. Then $\hat{z}_i \to \hat{z}$ as $i \to \infty$, and $\hat{z}$ satisfies $g(\hat{z}) = 0$.

<u>Proof</u>: First, it follows from Propositions (9) and (11) that $g(z_i) \to 0$ as $i \to \infty$, and hence, that all accumulation points $\hat{z}$ of $\{z_i\}_{i=0}^{\infty}$ must be in the set $\{z \mid g(z) = 0\}$. By assumption, this set consists of a finite number of points. Next, we must have

13. $\limsup \|z_{i+1} - z_i\| = 0$

either because $\delta \to 0$ as $i \to \infty$ or because $g(z_i) \to 0$ as $i \to \infty$ (since

$\|H_i^{-1}\| \le b$, only, is allowed in the construction $z_{i+1} = z_i - \beta^k H_i^{-1} g(z_i))$. Thus algorithm (3) satisfies the assumptions of theorem (1.3.66) in [7] which yields that the sequence $\{z_i\}_{i=0}^{\infty}$ converges.

14. Lemma: Suppose that assumptions (1(i)-1(iv)) are satisfied and that algorithm (3) has constructed a sequence $\{z_i\}_{i=0}^{\infty}$ with limit point $\hat{z}$. Then there exists an integer $N \ge 0$ and an $M > 0$ such that for all $i \ge N$,

15. $$\|G(z_i) - H_i\| \le M \sum_{j=i}^{i-n} \|z_j - \hat{z}\| \quad .$$

Proof: Since $z_i \to \hat{z}$ as $i \to \infty$, and $\varepsilon_i = \min\{\nu,\delta\}$ in Step 3 (where $\nu = \beta^k \|H_i^{-1} v\|$), it follows that there exists an integer $N \ge 0$ such that $\varepsilon_i < \rho/2$ and $\|z_i - \hat{z}\| < \rho/2$ for all $i \ge N$. Here ρ is as in (1(iii)). Now suppose that $i \ge N$ and (without loss of generality) that the jth column of H_i is $\frac{1}{\varepsilon_{i-k}} [g(z_{i-k}+\varepsilon_{i-k}e_j) - g(z_{i-k})]$, where $k \in \{0,1,2,\ldots,n-1\}$. Then, making use of (1(iii)) and the fact that $\varepsilon_{i-k} \le \|z_{i-k} - z_{i-k-1}\|$, we obtain that the magnitude of the difference between the jth columns of $G(z_i)$ and H_i is

16. $$\|G(z_i)e_j - \frac{1}{\varepsilon_{i-k}} [g(z_{i-k}+\varepsilon_{i-k}e_j) - g(z_{i-k})]\|$$

$$= \|\int_0^1 [G(z_i) - G(z_{i-k}+t\varepsilon_{i-k}e_j)]e_j dt\|$$

$$\le L \int_0^1 \|z_i - z_{i-k}-t\varepsilon_{i-k}e_j\| dt$$

$$\leq L \int_0^1 (\|z_i - z_{i-k}\| + t\,\|z_{i-k} - z_{i-k-1}\|)dt$$

$$\leq L(\|z_i-\hat{z}\| + \frac{3}{2}\,\|z_{i-k}-\hat{z}\| + \frac{1}{2}\,\|z_{i-k-1}-\hat{z}\|).$$

The existence of a constant M satisfying (15) now follows from (16) and the properties of norms on a Euclidean space.

17. __Lemma__: Suppose that all the assumptions (1) are satisfied, that $b \geq 2\|G(\hat{z})^{-1}\|$ for all $\hat{z} \in S$ and that the algorithm (3) has constructed a sequence $\{z_i\}_{i=0}^{\infty}$. Then there exists an integer $N \geq 0$ such that for all $i \geq N$, the test in step 7 is satisfied, (7) is satisfied with $k = 0$, and step 13 is not reached by the algorithm, i.e., for all $i \geq N$,

18. $$z_{i+1} = z_i - H_i^{-1}g(z_i).$$

__Proof__: First, since $g(\cdot)$ is twice continuously differentiable, we note that

$$\frac{\partial^2}{\partial z^2}\left(\frac{1}{2}\|g(z)\|^2\right) = G(z)^T G(z) + W(z)$$

where $W(z)$ is an nxn matrix, which is continuous in z, and which satisfies $W(\hat{z}) = 0$ for all $\hat{z} \in \{z | g(z) = 0\}$. Hence, if H_i^{-1} exists, then expanding $\|g(z_i-H_i^{-1}g(z_i))\|^2$ to second order terms according to the Taylor formula, we obtain,

19. $$\|g(z_i-H_i^{-1}g(z_i))\|^2 = \|g(z_i)\|^2 - 2\,\langle\, G(z_i)^T g(z_i),\ H_i^{-1}g(z_i)\,\rangle$$

$$+ 2\int_0^1 (1-s)(\langle H_i^{-1}g(z_i),\ G(z_i-sH_i^{-1}g(z_i))^T G(z_i-sH_i^{-1}g(z_i))H_i^{-1}g(z_i)\rangle$$

$$+ \langle H_i^{-1}g(z_i), W(z_i-sH_i^{-1}g(z_i))H_i^{-1}g(z_i)\rangle)ds$$

To simplify the expressions in (19), let $G_i(s) = G(z_i-sH_i^{-1}g(z_i))$ and $W_i(s) = W(z_i-sH_i^{-1}g(z_i))$. Hence (19) becomes

20. $$\|g(z_i-H_i^{-1}g(z_i))\|^2 = \|g(z_i)\|^2 \left(1 - 2\,\frac{\langle g(z_i),G(z_i)H_i^{-1}g(z_i)\rangle}{\|g(z_i)\|^2} + \frac{2}{\|g(z_i)\|^2}\int_0^1 (1-s)[\|G_i(s)H_i^{-1}g(z_i)\|^2 + \langle H_i^{-1}g(z_i),W_i(s)H_i^{-1}g(z_i)\rangle]ds\right)$$

Now, since $z_i \to \hat{z}$, it follows from (15) and the Perturbation Lemma (2.3.2) in [6] that there exists an integer N' such that for all $i \geq N'$, H_i^{-1} exists and is bounded. Consequently, for $s \in [0,1]$, $W_i(s) \to 0$ as $i \to \infty$, $G_i(s)H_i^{-1} \to I$ as $i \to \infty$, and $G(z_i)H_i^{-1} \to I$ as $i \to \infty$, where I is the nxn identity matrix. Therefore, given any $\alpha \in (0, \frac{1}{2})$, it follows from (20) that there exists an integer $N \geq N'$ such that for all $i \geq N$

21. $$\|g(z_i-H_i^{-1}g(z_i))\|^2 \leq (1-2\alpha)\|g(z_i)\|^2,$$

i.e. the test (7) is satisfied with $k = 0$ for all $i \geq N$. The fact that step 13 is not reached for $i \geq N$ is obvious. This completes our proof.

Since the multiplier of $\|g(z_i)\|^2$ in the right hand side of (20) goes to zero as $i \to \infty$, it is clear that $\|g(z_i)\|^2 \to 0$ superlinearly. However, we can make the following, stronger statement as well.

22. Theorem: Suppose that all the assumptions (1) hold and that algorithm (3) has constructed a sequence $\{z_i\}_{i=0}^{\infty}$ converging to the point $\hat{z} \in \{z \mid g(z) = 0\}$. Then

23. $$0 < \limsup_{i\to\infty} \|z_i - \hat{z}\|^{1/\tau_n^i} < 1,$$

where τ_n is the unique positive root of the equation $t^{n+1} - t^n - 1 = 0$ (i.e., the R-order of algorithm (3) is τ_n, where R-order is defined by

(9.2.5) in [6]).

Proof: Let N be such that lemmas (14) and (17) hold. Then, for all $i \geq N$ we have

24. $$z_{i+1} = z_i - H_i^{-1} g(z_i)$$

and hence, since $g(\hat{z}) = 0$, for all $i \geq N$,

25. $$\begin{aligned} \|z_{i+1} - \hat{z}\| &= \|(z_i-\hat{z}) - H_i^{-1}(g(z_i) - g(\hat{z}))\| \\ &\leq \left\| \int_0^1 (I-H_i^{-1}G(\hat{z}+s(z_i-\hat{z})))(z_i-\hat{z})ds \right\| \\ &\leq \int_0^1 \|H_i^{-1}(H_i-G(\hat{z}+s(z_i-\hat{z}))\| ds \; \|z_i-\hat{z}\| \\ &\leq \|H_i^{-1}\| \sup_{s \in [0,1]} \|H_i - G(\hat{z}+s(z_i-\hat{z}))\| \; \|z_i-\hat{z}\|. \end{aligned}$$

Since for $i \geq N$, $\|H_i^{-1}\| \leq b$, and by (1(iii)) $\|G(\hat{z}+s(z_i-\hat{z})) - G(z_i)\| \leq L(1-s)\|\|z_i-\hat{z}\|$, $s \in [0,1]$, it follows from (25) that for $i \geq N$,

26. $\|z_{i+1} - \hat{z}\| \le b\ (\|H_i - G(z_i)\| + L\ \|z_i - \hat{z}\|)\|z_i - \hat{z}\|.$

Finally, making use of (15), we obtain that there exist constants $\gamma_j \ge 0$, $j = 0,1,2,\dots,n-1$, such that

27. $$\|z_{i+1} - z\| \le \|z_i - \hat{z}\| \sum_{j=0}^{n} \gamma_j \|z_{i-j} - \hat{z}\|.$$

Our theorem now follows directly from theorem (9.2.9) in [6].

3. Applications

One of the more promising applications of the secant method described in Section 2 is in the solution of boundary value problems of the form

28. $$\frac{d}{dt} x(t) = h(x(t),t) \qquad t \in [t_0,t_f]$$

29. $$g_0(x(t_0)) = 0,\ g_f(x(t_f)) = 0$$

where g_0, g_f and $h : \mathbb{R}^n \times \mathbb{R}^1 \to \mathbb{R}^n$ are twice continuously differentiable in x, and $h(x,t)$, $\frac{\partial}{\partial x} h(x,t)$, $\frac{\partial^2}{\partial x^2} h(x,t)$ are all continuous (or at least piecewise continuous) in t. In addition, we assume that

30. $$g(z) \equiv \begin{pmatrix} g_0(z) \\ \\ g_f(x(t_f,z)) \end{pmatrix}$$

maps $\mathbb{R}^n$ into $\mathbb{R}^n$ and $G(z) = \frac{\partial}{\partial z} g(z)$ is nonsingular in a suitable ball in $\mathbb{R}^n$. In (30), $x(t_f,z)$ denotes the solution of (28) at $t = t_f$, obtained from the initial condition $x(t_0) = z$. Since $g(z)$ and $G(z)$ are quite expensive to calculate, it is clear that (28), (29) represents a class of problems in which a good secant method could do considerably better than Newton's method.

In our nemerical experiments with reasonably well behaved problems of the form (28), (29), we have initialized δ at $\delta = 0.2 \max_j |z_0^j|$ and we have found that the algorithm would pass the test (7) after a very small number of iterations (often, $i < 3$). We have also found, as expected, that the total computing time needed to reach $\|g(z_i)\| \leq 10^{-6}$ was much smaller with our secant method than with Newton's method.

Conclusion

A limited amount of numerical experimentation indicates that algorithm (3) is a highly efficient method for solving equations in several variables. In application to boundary value problems, it is subject to the same difficulties as the Newton method, whenever these difficulties are caused by the ill-conditioning of the Jacobian matrix $G(z)$. In the case of Newton's method, it is sometimes possible to reduce this ill-conditioning by means of a nonlinear transformation such as the one due to Abramov [1]. It remains to be seen whether it is possible to graft Abramov's procedure onto a secant method such as ours without destroying its efficiency and without creating unreasonable storage demands.

Finally, it should be pointed out that when used for solving boundary value problems, to obtain greater efficiency, the method should be modified so as to determine adaptically the required integration precision at each iteration. A general theory indicating how this is to be performed is given in Appendix A of [7], while two specific examples can be found in [5] and [8].

Acknowledgement: The author wishes to thank Mr. J. Raphel for programming and testing the algorithm in this paper, and Mr. H. Mukai for useful discussions.

REFERENCES

[1] A. A. Abramov, On the transfer of boundary conditions for systems of ordinary linear differential equations, Zh. Vychisl. Mat. Mat. Fiz. 1 (3), 542-545(1961).

[2] L. Armijo, Minimization of functions having continuous partial derivatives, Pacific J. Math. 16, 1-3(1966).

[3] N. V. Banitchouk, V. M. Petrov, and R. L. Chernousko, Numerical Solution of Problems with variational limits by the method of local variations, Zh. Vychisl. Mat. Mat. Fiz. 6 (6), 947-961(1966).

[4] J. Barnes, An Algorithm for solving nonlinear equations based on the secant method, Comp. J. 8, 66-72,(1965).

[5] R. Klessig and E. Polak, An adaptive algorithm for unconstrained optimization with an application to optimal control, Univ. of Calif. ERL Memo M297, Feb. 22, 1971. SIAM J. Control, in press.

[6] J. M. Ortega, and W. C. Rheinboldt, Iterative solution of nonlinear equations in several variables, Academic Press, N. Y.,1970.

[7] E. Polak, Computational Methods in Optimization: A Unified Approach, Academic Press, New York, 1971.

[8] E. Polak, On a class of numerical methods with an adaptive integration subprocedure for optimal control problems, Proc. 4th IFIP Colloqium on Optimization, Santa Monica, Calif., Oct. 19-22, 1971.

[9] P. Wolfe, The secant method for simultaneous nonlinear equations, Comm. ACM, vol. 2, pp. 12-14, 1959.