A MODIFIED SECANT METHOD FOR UNCONSTRAINED MINIMIZATION

by

E. Polak

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

1. Introduction

The one variable secant method, for the solution of equations, has been known for a very long time as being computationally more efficient than Newton's method. Among the extensions of this method to n-dimensional problems, those proposed by Wolfe [8] and Barnes [2], are among the most interesting ones, because, when they do converge, they are computationally considerably more efficient than Newton's method (see, for example, the discussion in [5]). However, as can be seen from the counter examples quoted in [5], these methods may not converge. More recently, Ritter [7] proposed a new algorithm for function minimization, combining Goldstein's gradient method [3] with a secant type method, which contains an angle test (between the direction of descent and the gradient) to ensure convergence. This angle test depends on a parameter that may be quite difficult to select in advance. A bad selection results in the algorithm staying in the gradient mode most, if not all the time.

Research sponsored by the National Aeronautics and Space Administration under Grant NGL-05-003-16, the National Science Foundation Grant GK-10656X2, and the Joint Services Electronics Program, Contract F-44620-71-C-0087.

In this paper we present a new gradient-secant algorithm for unconstrained optimization problems of the form min $\{f(z) \mid z \in \mathbb{R}^n\}$. It differs from Ritter's method both in the fact that it does not use an angle test and in the manner in which it updates the approximate hessian. Roughly speaking, in solving a problem, this algorithm uses Armijo gradient method iterations [1] until it reaches a region where the Newton method is more efficient than the gradient method. Then it switches over to a secant form of operation. Under the assumption that f is continuously differentiable, we have shown that any accumulation point z, of a sequence constructed by this algorithm, must be stationary. Under the stronger hypothesis that f is twice continuously differentiable and strictly convex, we were able to show that any sequence $\{z_i\}_{i=0}^{\infty}$ constructed by our algorithm converges superlinearly to the unique minimizer $\hat{z}$ of $f(\cdot)$, with rate τ^n, where τ^n is the unique positive root of $t^{n+1} - t^n - 1 = 0$, i.e. that for some $\theta \in (0,1)$ and some $R \in (0,\infty)$, $\| z_i - \hat{z} \| \leq R\, \theta^{\tau_n^i}$, $i=0, 1, 2, \ldots$. Both theoretical considerations and our computational experiments indicate that this new algorithm is considerably faster than the Newton method, and Lootsma [4] reports that on many problems Newton's method is superior to a number of conjugate direction and quasi-Newton methods. It is therefore not unrealistic to hope that, as experience with the new method accumulates, it will emerge as one of the most efficient methods for the solution of certain classes of unconstrained optimization problems.

2. The Secant Method.

Consider the problem

1. $\min\{f(z) \mid z \in \mathbb{R}^n\}$

To begin, we shall make only the following minimal assumptions.

2. Assumptions: (a) f is continuously differentiable and, (b) f is bounded from below. ¤

Throughout this paper, when we say that an algorithm is convergent, we mean that every limit point $\hat{z}$ of a sequence it constructs in solving (1) satisfies $f(\hat{z}) = 0$.

The assumptions (2) will suffice to prove that the algorithm we are about to state is convergent. We shall later show, under stronger assumptions, that it converges superlinearly and establish a bound on its rate of convergence.

We shall use the notation

3. $g(z) \equiv \nabla f(z),\ z \in \mathbb{R}^n.$

4. Algorithm:

Data: $\delta > 0$, $\alpha \in (0, \frac{1}{6})$, $\beta \in (0,1)$, $b > 0$ (large)†, $\ell \geq 2$, $z_0 \in \mathbb{R}^n$, H a symmetric positive definite $n \times n$ matrix, $e_j = j^{th}$ column of $n \times n$ unit matrix, $j = 1, 2, .., n$.

Step 0: Set $i = 0$, $j = 0$, $p = 0$, $\nu_0 = \delta$, $\bar{H} = H$. Compute $g(z_0)$ and set $\gamma_0 = \nu_0 = \| g(z_0) \|^2$.

Step 1: Compute $g(z_i)$. Stop if $g(z_i) = 0$.

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

† The purpose of the constant b is to make the algorithm use a steepest descent step whenever H_i, the current approximation to the hessian of $f(\cdot)$ is "too close" to being singular. A lower bound on b is $b \geq 2 \| H(\hat{z})^{-1} \|$ for all $\hat{z}$ which are local minimizers of $f(\cdot)$. In practice, setting $b = \infty$ does not appear to destroy the convergence of the algorithm.

go to step 3.

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

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

Step 5: Replace $\bar{h}_j$, the $j^{\underline{th}}$ column of $\bar{H}$, by

5. $$\Delta_i = \frac{1}{\varepsilon_i} [g(z_i + \varepsilon_i e_j) - g(z_i)]$$

to obtain a new matrix $\bar{H}$, and set $H_i = \bar{H}$.†

Step 6: If $\| g(z_i) \| \leq \gamma_p$, go the step 7; else set $W = z_i$ go to step 15.

Step 7: If H_i^{-1} exists and $\| H_i^{-1} \| \leq b$, compute

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

and go to step 8; else set $w = z_i$ and go to step 15.

Step 8: If $\langle v_i, g(z_i) \rangle < 0$, go to step 9; else set $w = z_i$ and go to step 15.

Step 9: Set $k = 0$.

Step 10: Compute $f(z_i - \beta^k v_i)$.

Step 11: If

7. $$f(z_i - \beta^k v_i) - f(z_i) < 0$$

go to step 13; else go to step 12.

Step 12: If $k < \ell$, set $k = k + 1$ and go to step 10; else go to step 15.

Step 13: Compute $g(z_i - \beta^k v_i)$. If $g(z_i - \beta^k v_i) = 0$, set $z_{i+1} = z_i - \beta^k v_i$ and stop.

Step 14: If

8. $$\| g(z_i - \beta^k v_i) \|^2 \leq (1 - 2\beta^\ell \alpha) \| g(z_i) \|^2,$$

set $z_{i+1} = z_i - \beta^k v_i$, set $\gamma_{p+1} = \| g(z_{i+1}) \|^2$, set $p = p+1$, set $i = i+1$

† Note that since H_{new} differs from H_{old} in only one column, $H_i^{-1} = H_{new}^{-1}$ can be obtained from H_{old}^{-1} by means of the standard updating formula.

and go to step 2; else set $w = z_i - \beta^k \nu_i$ and go to step 15.

Step 15: Compute the smallest integer $s_i \geq 0$ such that

9. $$f(z_i - \beta^{s_i} g(z_i)) - f(z_i) \leq - \beta^{s_i} \alpha \| g(z_i) \|^2$$

and set $y = z_i - \beta^{s_i} g(z_i)$.

Step 16: If $f(y) < f(w)$, set $z_{i+1} = y$, set $i = i+1$, and go to step 1; else set $z_{i+1} = w$, set $i = i+1$ and go to step 1. ¤

Since when $g(z_i) \neq 0$, one can always find a finite s_i such that (9) is satisfied, algorithm (4) is obviously well defined.

10. Theorem: Suppose that the assumptions (2) are satisfied and that algorithm (4) has constructed an infinite sequence $\{z_i\}_{i=0}^{\infty}$. Then every limit point z^* of $\{z_i\}_{i=0}^{\infty}$ satisfies $g(z^*) = 0$.

Proof: Suppose that z^* is a limit point of $\{z_i\}$, that $g(z^*) \neq 0$ and that $z_i \to z^*$ for $i \in K$, with K an infinite subset of the positive integers. Now there are two possibilities.

(i) There exists an infinite subset $K' \subset K$ such that for all $i \in K'$, either

11. $$z_{i+1} = z_i - \beta^{s_i} g(z_i)$$

or

$$z_{i+1} = z_i - \beta^k v_i$$

12.

and

13. $$f(z_i - \beta^k v_i) < f(z_i - \beta^{s_i} g(z_i)).$$

Since $g(z^*) \neq 0$ and $z_i \to z^*$ for $i \in K'$, it follows from the discussion in sec. 2.1 of [6] (see theorem 22 and algorithm 35) that there exists a $\delta(z^*) < 0$ and an integer N such that

14. $f(z_{i+1}) - f(z_i) \leq \delta(z^*) < 0$ for all $i \geq N$, $i \in K'$.

But K' is an infinite subset, and $\{f(z_i)\}_{i=0}^{\infty}$ is a monotonically decreasing sequence, hence (14) contradicts the assumption that f(z) is bounded from below. Thus, if (i) holds, then $g(z^*) = 0$.

The second possibility is

(ii) There exists an infinite subset $K'' \subset K$ such that

15. $z_{i+1} = z_i - \beta^k v_i$ for all $i \in K''$

and

16. $\| g(z_{i+1}) \|^2 \leq (1-2\beta^{\ell}\alpha) \| g(z_i) \|^2$ for all $i \in K''$

In this case the sequence $\{\gamma_p\}$ is infinite, monotonically decreasing and bounded from below by zero. Hence $\gamma_p \to \gamma^* \geq 0$. Now, since whenever (15) and (16) take place, $\gamma_{p+1} = \| g(z_{i+1}) \|^2$, for some integer p, and $\gamma_p \geq \| g(z_i) \|^2$, $\gamma_{p+1} - \gamma_p \leq \gamma_{p+1} - \| g(z_i) \|^2 \leq - 2\beta^{\ell}\alpha \| g(z_i) \|^2$. Hence, since $z_i \to z^*$ for $i \in K''$, and since $g(\cdot)$ is continuous by assumption (2), there exists an infinite subset K''' of the positive integers such that

17. $\gamma_{p+1} - \gamma_p \leq -\beta^{\ell}\alpha \| g(z^*) \|^2$ for all $p \in K''$.

But (17) contradicts the fact that $\gamma_p \to \gamma^* \geq 0$.

Hence we must have $g(z^*) = 0$. ¤

The following result is a direct consequence of theorem (10).

18. <u>Corollary:</u> Algorithm (4) is convergent whenever problem (1) satisfies

the assumptions (2). ⌑

The following corollary can be deduced from theorem (10), which implies that $g(z_i) \to 0$ as $i \to \infty$ and hence that $z_{i+1} - z_i \to 0$ as $i \to \infty$, and theorem (1.3.66) in [6].

19. Corollary: Suppose that the sequence $\{z_i\}_{i=0}^{\infty}$ described in theorem (10) is compact and that the function $f(\cdot)$ has only a finite number of stationary points, then there exists a $z^* \in \mathbb{R}^n$ such that $z_i \to z^*$ and $g(z^*) = 0$. ⌑

We are now ready to establish the rate of convergence of algorithm (4). For this purpose we shall need to assume the following.

20. Assumption: The function f is (a) three times continuously differentiable and, (b) strictly convex. ⌑

Note that under assumption (20), the level sets of $f(\cdot)$ are compact and there exists only one point $\hat{z}$ (the minimizer of $f(z)$ over $\| ^n$), which satisfies $g(\hat{z}) = 0$. Hence, by theorem (10) and corollary (19), whenever assumption (20) is satisfied, any sequence $\{z_i\}_{a=0}^{\infty}$ constructed by algorithm (4) converges to the unique minimizer $\hat{z}$ of $f(\cdot)$.

21. Lemma: Suppose that assumption (20) is satisfied and that algorithm (4) has constructed an infinite sequence $\{z_i\}_{i=0}^{\infty}$ converging to $\hat{z}$, the minimizer of $f(\cdot)$, Then there exists $0 \leq M < \infty$ such that

22. $$\| H(z_i) - H_i \| \leq M \sum_{j=i}^{i-n} \| z_j - \hat{z} \| \text{ for } i = 0, 1, 2, \ldots,$$

where H_i is as defined in the algorithm and

23. $H(z) \equiv \frac{\partial^2 f(z)}{\partial z^2}, \ z \in \mathbb{R}^n.$

Proof: Since (20) (b) is satisfied, the level set $C(z_0) = \{z \mid f(z) \leq f(z_0)\}$ is compact and convex and hence, since (20) (a) is satisfied, there exists a Lipshitz constant $L < \infty$ such that for all $x, y \in C(z_0)$,

24. $\| H(x) - H(y) \| \leq L \| x - y \|$

(Note that $\{z_i\}_{i=0}^{\infty}$ is contained in $C(z_0)$). Now, without loss of generality, suppose that the jth column of $H_i (j \in \{1, 2, \ldots, n\})$ is

25. $\frac{1}{\varepsilon_{i-k}} [g(z_{i-k} + \varepsilon_{i-k} e_j) - g(z_{i-k})]$, where $k \in \{0, 1, 2, .., n-1\}$.

Then, making use of (24), of the mean value theorem, and the fact that $\varepsilon_{i-k} \leq \| z_{i-k} - z_{i-k-1} \|$ by construction, we obtain that the magnitude of the difference between the j^{th} columns of $H(z_i)$ and H_i satisfies

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

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

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

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

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

The existence of a constant M satisfying (22) now follows from (26) and the triangle inequality for norms, used in conjunction with the addition and subtraction of terms in the right hand side of (6). ∎

27. Lemma: Suppose that assumption (20) is satisfied, that $b \geq 2 \| H(\hat{z})^{-1} \|$ and that the algorithm (4) has constructed a sequence $\{z_i\}_{i=0}^{\infty}$. Then there exists an integer N such that for all $i \geq N$,

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

Proof: First, since $z_i \to \hat{z}$, the global minimizer of $f(\cdot)$, and (22) holds, it follows from the perturbation Lemma (2.3.2) in [5] that there exists an integer N' such that for all $i \geq N'$ H_i^{-1} exists and is positive definite and $\| H_i^{-1} \| \leq b$. Hence, for all $i \geq N'$, the test in step 8 of the algorithm, i.e., $\langle v_i, g(z_i) \rangle = - \langle H_i^{-1} g(z_i), g(z_i) \rangle < 0$, is satisfied for all $i \geq N'$ and hence the computation proceeds to step 9.

Next, applying the second order Taylor expansion, we obtain (with $v_i = H_i^{-1} g(z_i)$)

28. $$f(z_i - H_i^{-1} g(z_i)) - f(z_i) =$$

$$- \langle g(z_i), H_i^{-1} g(z_i) \rangle + \int_0^1 (1-t) \langle H_i^{-1} g(z_i), H(z_i - tv_i) H_i^{-1} g(z_i) \rangle dt$$

$$= - \langle g(z_i), H_i^{-1} g(z_i) \rangle + \int_0^1 (1-t) [\langle H_i^{-1} g(z_i), H(\hat{z}) H_i^{-1} g(z_i) \rangle$$

$$+ \langle H_i^{-1} g(z_i), (H(z_i - t v_i) - H(\hat{z}))\, H_i^{-1} g(z_i) \rangle]\, dt$$

Hence, since $v_i = H_i^{-1} g(z_i) \to 0$ as $i \to \infty$ (because $H_i^{-1} \to H(\hat{z})^{-1}$ and $g(z_i) \to 0$ causing $H(z_i - v_i) - H(\hat{z}) \to 0$ uniformly for $t \in [0,1]$ as $i \to \infty$), and since $H(\hat{z})\, H_i^{-1} \to I$, the unit matrix, as $i \to \infty$, it follows from (28) that there is an integer $N'' \geq N'$ such that

29. $$f(z_i - H_i^{-1} g(z_i)) - f(z_i) < 0 \text{ for all } i \geq N'',$$

i.e. the test in step 11 of the algorithm is satisfied with $k = 0$ for all $i \geq N''$.

Finally, consider $\frac{1}{2} \| g(z) \|^2$. Since f is three times continuously differentiable,

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

where $W(z)$ is a continuous nxn matrix which satisfies $W(\hat{z}) = 0$. Hence, for all $i \geq N''$, expanding $\| g(z_i - H_i^{-1} g(z_i)) \|^2$ to second order terms according to the Taylor formula, we obtain

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

$$+ 2 \int_0^1 (1-t)\, [\langle H_i^{-1} g(z_i), H(z_i - t v_i)^T H(z_i - t v_i)\, H_i^{-1} g(z_i) \rangle$$

$$+ \langle H_i^{-1} g(z_i), W(z_i - t v_i)\, H_i^{-1} g(z_i) \rangle]dt,$$

Where $v_i = H_i^{-1} g(z_i)$. Setting $\bar{H}_i(t) = H(z_i - t v_i)$ and $\bar{W}_i(t) = W(z_i - t v_i)$, $t \in [0,1]$, (31) yields

32. $$\| g(z_i - v_i) \|^2 = \| g(z_i) \|^2 \left[1 - 2 \frac{\langle g(z_i), H(z_i) H_i^{-1} g(z_i) \rangle}{\| g(z_i) \|^2} + \frac{2}{\| g(z_i) \|^2} \int_0^1 (1-t) \| \bar{H}_i(t) H_i^{-1} g(z_i) \|^2 + \langle H_i^{-1} g(z_i), \bar{W}_i(t) H_i^{-1} g(z_i) \rangle \right] dt$$

Since $v_i \to 0$ as $i \to \infty$ and $z_i \to \hat{z}$ as $i \to \infty$, $\bar{W}_i(t) \to 0$ as $i \to \infty$, uniformly in $t \in [0,1]$ and similarly, $\bar{H}_i(t) H_i^{-1} \to I$ as $i \to \infty$, uniformly in $t \in [0,1]$. Hence the term in the right hand side of (32), multiplied by $\| g(z_i) \|^2$ tends to zero as $i \to \infty$ and therefore there exists an integer $N''' \geq N''$ such that the test (8) is satisfied for $k = 0$ for all $i \geq N'''$.

Now, since $g(z_i) \to 0$ as $i \to \infty$, there exists an integer $N \geq N'''$ at which the test in step 6, viz., $\| g(z_i) \|^2 \leq \gamma_p$ will be satisfied. Then, for all $i \geq N$, $z_{i+1} = z_i - H_i^{-1} g(z_i)$, which completes our proof. ¤

33. Theorem: Suppose that assumption (20) is satisfied and that algorithm (4) has constructed a sequence $\{z_i\}_{i=0}^{\infty}$. Then

34. $$0 < \limsup_{i \to \infty} \| z_i - \hat{z} \|^{1/\tau_n^i} < 1,$$

where τ_n is the unique positive root of the equation $t^{n+1} - t^n - 1 = 0$ and $\hat{z}$ is the unique minimizer of $f(\cdot)$ (i.e., the R-order of algorithm (4) is τ_n, where R-order is defined by (9.2.5) in [5]).

Proof: Let N be an integer such that for all $i \geq N$

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

By Lemma (27), such an N exists. Then since $H(\cdot)$ is Lipshitz continuous on $C(z_0)$, for all $i \geq N$ (since $g(\hat{z}) = 0$)

36. $\| z_{i+1} - \hat{z} \| = \| (z_i - \hat{z}) - H_i^{-1}(g(z_i) - g(\hat{z})) \|$

$$\leq \| \int_0^1 (I - H_i^{-1} H(\hat{z} + t(z_i - \hat{z}))(z_i - \hat{z})\, dt \|$$

$$\leq \int_0^1 \| H_i^{-1} \| \| (H_i - H(\hat{z} + t(z_i - \hat{z})) \| \| z_i - \hat{z} \| dt$$

$$\leq \| H_i^{-1} \| \int_0^1 ([\| H_i - H(z_i) \| + \| H(z_i) - H(\hat{z} + t(z_i - \hat{z})) \|] \| z_i - \hat{z} \|)\, dt$$

$$\leq \| H_i^{-1} \| \int_0^1 [\| H_i - H(z_i) \| + Lt \| z_i - \hat{z} \|] \| z_i - \hat{z} \| dt,$$

where L is the Lipshitz constant for $H(\cdot)$ on $C(z_0)$. Now making use of Lemma (22) and the fact that $\| H_i^{-1} \|$ is bounded for $i \geq N$ (since $H_i^{-1} \to H(\hat{z})^{-1}$), we conclude from (36) that there exists constants $\lambda_j \geq 0$, $j = 0, 1, 2, .., n-1$, such that for all $i \geq N$

37. $\| z_{i+1} - \hat{z} \| \leq \| z_i - \hat{z} \| \sum_{j=0}^{n} \lambda_j \| z_{i-j} - \hat{z} \|.$

The desired result now follows from (7) and theorem (9.2.9) in [5]. ∎

38. <u>Corollary:</u> Under the conditions in theorem (33), any sequence $\{z_i\}_{i=0}^{\infty}$ constructed by algorithm (4) satisfies

39. $$\limsup_{i\to\infty} \| g(z_i) \|^{1/\tau_n^i} < 1.$$

40. $$\limsup_{i\to\infty} [f(z_i) - f(\hat{z})]^{1/\tau_n^i} < 1$$

<u>Proof:</u> Since $g(\hat{z}) = 0$,

41. $$\| g(z_i) \| = \| \int_0^1 H(z_i+t(z_i-\hat{z}))\,(z_i-\hat{z})\,dt \|$$

$$\leq [\int_0^1 \| H(z_i+t(z_i-\hat{z})) \| dt] \| z_i - \hat{z} \|$$

$$\leq Q \| z_i-\hat{z} \|$$

where $Q = \sup \{\| H(z) \| \mid z \in C(z_0)\}$. Relation (39) now follows from (41), (34) and the fact that $Q^{1/\tau_n^i} \to 1$ as $i \to \infty$.

Next, again since $g(\hat{z}) = 0$,

42. $$f(z_i) - f(\hat{z}) = \int_0^1 (1-t)\langle (z_i-\hat{z}), H(\hat{z}+t(z_i-\hat{z}))\ (z_i-\hat{z})\rangle\, dt$$

$$\leq \int_0^1 (1-t)\ \bar{Q}\ \| z_i - \hat{z} \|^2\, dt$$

$$= \frac{1}{2}\bar{Q}\ \| z_i - \hat{z} \|^2$$

where $\bar{Q}$ is an upper bound on the eigenvalues of $H(z)$ for $z \in C(z_0)$.

Relation (40) now follows from (42), (34) and the fact that $(\frac{1}{2}\bar{Q})^{1/2\tau_n^i} \to 1$ as $i \to \infty$. This completes our proof. ¤

We note that the only time we made use of the assumption that $f(\cdot)$ was three times continuously differentiable and strictly convex was in the proof of lemma (27). At this point is is easy to show that lemma (27) can also be proved under the weaker assumption that $f(\cdot)$ is only twice continuously differentiable strictly convex and (24) holds. Thus suppose that $\{z_i\}_{i=0}^{\infty}$ is any sequence such that $z_i \to \hat{z}$ as $i \to \infty$ and that $y_{i+1} = z_i - H_i^{-1}\, g(z_i)$ for all $i \geq \bar{N}$, where $\bar{N}$ is such that H_i^{-1} exists for all $i \geq \bar{N}$. Then (37) applies and yields

43. $$\| y_{i+1} - \hat{z} \| \leq \| z_i - \hat{z} \| \sum_{j=0}^{n} \| z_{i-j} - \hat{z} \|, \text{ for all } i \geq \bar{N}$$

Now, by the mean value theorem, since $g(\hat{z}) = 0$,

44. $$\| g(y_{i+1}) \| \leq \int_0^1 \| H(y_{i+1} + s(y_{i+1} - \hat{z})\| \, ds\ \| y_{i+1} - \hat{z} \|$$

$$\leq Q \| y_{i+1}-\hat{z} \| \cdot$$

$$\leq Q \| z_i-\hat{z} \| \sum_{j=0}^{n} \| z_{i-j}-\hat{z} \|,$$

where Q is as in (41).

Now,

45. $$\| z_i-\hat{z} \| \| g(z_i) \| \geq \langle z_i-\hat{z}, g(z_i)\rangle$$

$$= \int_0^1 \langle z_i-\hat{z}, H(z_i+s (z_i-\hat{z}) (z_i-\hat{z})\rangle ds$$

$$\geq \bar{Q} \| z_i-\hat{z} \|^2 ,$$

where $\bar{Q}$ is as in (42). Hence for all $i \geq \bar{N}$

46. $$\| g(z_i) \| \geq \bar{Q} \| z_i-\hat{z} \|$$

and therefore (44) yields

47. $$\| g(y_{i+1}) \| \leq \|g(z_i) \| \{\frac{1}{\bar{Q}} \sum_{j=0}^{n} \| z_{i-j}-\hat{z} \|\} \text{ for all } i \geq \bar{N}$$

Since $\sum_{j=0}^{n} \| z_{i-j}-\hat{z} \| \to 0$ as $i \to \infty$, we conclude that there exists an integer $\tilde{N} \geq \bar{N}$ such that

$$\| g(y_{i+1}) \|^2 \leq (1-2\alpha\beta^{\ell}) \| g(z_i) \|^2 \text{ for all } i \geq \tilde{N}$$

and hence that lemma (27) holds under the weaker assumption that f is only twice continuously differentiable and strictly convex.

Conclusion

We have presented in this paper an efficient method for unconstrained minimization. It should be clear from the development that the assumptions used to establish rate of convergence can be relaxed from a global statement to a local one, i.e. as holding in a convex neighborhood of a local minimum. It is also clear that one can construct several other variants of the algorithm as, for example, by substituting a conjugate directions method for the gradient method in the algorithm. In some applications these alternative, more complex versions may be preferred over the simplest one presented in this paper. As long as one substitutes for the Armijo gradient method any other convergent minimization method, the convergence and rate of convergence theorems, presented in this paper, remain valid.

REFERENCES

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

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

[3] A. A. Goldstein, "Constructive Real Analysis, Harper and Rowe, New York, 1967, Ch. 1, Section D.

[4] F. A. Lootsma, "Penalty function performance of several unconstrained-minimization techniques," Philips Res. Repts. 27, 358-385 (1972).

[5] J. M. Ortega and W. C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1971.

[6] E. Polak, Computational Methods in Optimization, Academic Press, New York, 1971.

[7] K. Ritter, "A Superlinearly convergent method for unconstrained minimization," in Nonlinear Programming, J. B. Rosen, O. L. Mangasarian and K. Ritter, Eds., Academic Press, New York 1970, pp. 177-206.

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