THE UNIVERSITY
OF WISCONSIN
madison, wisconsin

THE RATE OF CONVERGENCE OF THE RESET DAVIDON VARIABLE METRIC METHOD

Garth P. McCormick

MRC Technical Summary Report # 1012
October 1969

UNITED STATES ARMY

MATHEMATICS RESEARCH CENTER

THE UNIVERSITY OF WISCONSIN

MATHEMATICS RESEARCH CENTER, UNITED STATES ARMY

Contract No.: DA-31-124-ARO-D-462

THE RATE OF CONVERGENCE OF THE RESET DAVIDON VARIABLE METRIC METHOD

Garth P. McCormick

MRC Technical Summary Report # 1012

October 1969

Madison, Wisconsin 53706

ABSTRACT

The rate of convergence of the Reset Davidon Variable Metric Method for minimizing an unconstrained function of n variables is considered. If the limit point of the sequence of points generated by the method is a stationary point with a positive definite Hessian, the rate of convergence is superlinear with respect to cycles of n points. With an additional Lipschitz assumption the rate of convergence is shown to be at least quadratic for subsets of cycles of n points.

THE RATE OF CONVERGENCE OF THE RESET DAVIDON VARIABLE METRIC METHOD

Garth P. McCormick

Introduction

The Variable Metric Method proposed by Davidon [1] for solving the unconstrained minimization problem

$$\text{minimize} \quad f(x) \tag{1}$$

where $x = (x_1, \ldots, x_n)'$ and $f \in C^1$ is summarized as follows.

Davidon Variable Metric Method (DVMM)

STEP 0: Let $H^0 \equiv$ some arbitrary symmetric positive definite matrix ($z'H^0z > 0$ for all $z \neq 0$), x^0 be some arbitrary initial starting point. Set $s^0 = -H^0g^0$. Proceed as in the general step k+1 at equation (8).

STEP k + 1: At point x^{k+1}, $(k = 0, \ldots)$ let

$$\sigma^k \equiv x^{k+1} - x^k, \tag{2}$$

$$y^k \equiv g^{k+1} - g^k \tag{3}$$

$$H^{k+1} \equiv H^k - H^k y^k \alpha^k (\gamma^k)' H^k + \sigma^k \beta^k (\sigma^k)', \tag{4}$$

Sponsored by the Mathematics Research Center, United States Army, Madison, Wisconsin, under Contract No.: DA-31-124-ARO-D-462.

where

$$\alpha^k \equiv ((y^k)' H^k y^k)^{-1} \tag{5}$$

$$\beta^k \equiv ((\sigma^k)' y^k)^{-1}, \tag{6}$$

where

$g^k \equiv \nabla f(x^k)$, the vector of first partial derivatives of $f(x)$ at x^k. Let the direction vector at x^{k+1} be given by

$$s^{k+1} \equiv -H^{k+1} g^{k+1} . \tag{7}$$

Let t^{k+1} be the smallest local minimizing point of the one dimensional programming problem:

$$\underset{t \geq 0}{\text{minimize}} \quad f(x^{k+1} + s^{k+1} t) . \tag{8}$$

Let

$$x^{k+2} \equiv x^{k+1} + s^{k+1} t^{k+1} . \tag{9}$$

Repeat the step for x^{k+2}. If at any iteration k, $g^k = 0$, the procedure ceases.

Fletcher and Powell [2] showed that the above DVMM minimizes $f(x)$ where $f(x)$ is a positive definite quadratic form in n steps or fewer.

No one has been able to show[1] that for a general function $f(x)$, limit points of $\{x^k\}$ are stationary points (points where the vector of first derivatives vanishes). Also given that a limit point x^* is a

[1] The author has just received a copy of a paper by Powell [5] who proves that the DVMM converges to a stationary point if $f(x)$ is a twice differentiable convex function whose Hessian matrix has eigenvalues bounded below away from zero. Powell also shows that the rate of convergence is superlinear every step for this case if an additional Lipschitz condition is placed on the second derivatives of $f(x)$.

stationary point, no one has been able to ascertain the rate at which the sequence $\{x^k\}$ converges to x^*. Below is given a simple revision of the DVMM called the Reset Davidon Variable Metric Method (RDVMM). For this revised algorithm a statement of convergence is given in Theorem 1, and proofs of the rate of convergence in Theorems 2 and 3. In the discussion following Theorem 3 the differences between the convergence and rate of convergence of the RDVMM and the original DVMM are given.

Reset Davidon Variable Metric Method (RDVMM)

In step $k+1$, if $(k+1) \equiv 0 \mod(n+1)$, then set

$$H^{k+1} = H^0 \text{ (a symmetric positive definite matrix).} \tag{10}$$

Then equations (4) - (6) are bypassed.

It is useful here to state without proof some properties which apply to both algorithms.

For all k, H^k is a positive definite matrix. (11)

The direction vector $s^k = 0$ if and only if $g^k = 0$. (12)

Unless $s^k = 0$, $f(x^{k+1}) < f(x^k)$. (13)

Because of (8), and the fact that $f \in C^1$,

$$(g^{k+1})'s^k = 0, \tag{14}$$

i.e. the gradient of each point in the sequence is orthogonal to the previous direction.

If $f \in C^2$, using Taylor's theorem, for $j = 1, \ldots, n$,

$$y_j^k = \sum_{i=1}^{n} (\partial^2 f(\eta^{k,j})/\partial x_i \partial x_j)\sigma_i^k, \tag{15}$$

where each $\eta^{k,j}$ is a convex combination of x^k and x^{k+1}. It is convenient to define a matrix $\hat{G}(\eta^k)$ whose i,j^{th} element is

$$\partial^2 f(\eta^{k,i})/\partial x_j \partial x_i . \tag{16}$$

Then the equations expressed by (15) can be summarized as

$$y^k = \hat{G}(\eta^k)\sigma^k . \tag{17}$$

Furthermore, because of (8), for $j = 1, \ldots, n$

$$f(x^{k+1}) \leq f(\eta^{k,j}) \leq f(x^k) \tag{18}$$

where either equality holds in (18) only if

$$x^{k+1} = \eta^{k,j} \quad \text{or} \quad x^k = \eta^{k,j} .$$

The formula for t^k, using (2), (7), (9), (14) and (17) is

$$t^k = (g^k)'H^k g^k/(g^k)'H^k\hat{G}(\eta^k)H^k g^k . \tag{19}$$

For the RDVMM it is useful to divide the sequence (x^k) into overlapping groups of $(n + 2)$ points. The subscript c will be added to indicate the group $(c = 1, \ldots)$, and the superscript k will be used to indicate the order within the group $(k = 0, 1, \ldots, n+1)$.

The last point of each group is the same as the first point of the next group. The sequence then looks like

$$\begin{array}{l} x_1^0, \ldots, x_1^{n+1} \\ \qquad x_2^0, \ldots, x_2^{n+1} \\ \qquad\qquad \cdot \\ \qquad\qquad\quad \cdot \\ \qquad\qquad\qquad x_c^{n+1} \\ \qquad\qquad\qquad x_{c+1}^0, \ldots \quad . \end{array} \tag{20}$$

It is now possible to state a convergence theorem about the RDVMM.

Theorem 1 [Convergence of RDVMM].

Assume $f \in C^1$. If x^* is a limit point of $\{x_c^0\}$ generated by the RDVMM, then

$$g^* = \nabla f(x^*) = 0. \tag{21}$$

Proof: The proof follows from (10), (7), (8), (9), and the arguments used in [4, Theorem 1]. Q.E.D.

Well-known necessary conditions that x^* be a local unconstrained minimum of $f(x)$ are that it be a stationary point, (i.e. that (21) hold) and that (if $f \in C^2$),

$$G^* = \nabla^2 f(x^*)$$

be a positive semi-definite matrix.

Sufficient conditions that x^* an on isolated local unconstrained minimum are that (21) hold and that

$$G^* \text{ be a positive definite matrix.} \qquad (22)$$

It is for a limit point satisfying (21) and (22) that rate of convergence can be determined.

Theorem 2 [Superlinear Convergence].

If: (1) $f \in C^2$,

(2) the RDVMM is applied to problem (1),

then: (a) every limit point x^* of $\{x_c^0\}$ is a stationary point.

If, in addition,

(3) G^* is a positive definite matrix,

then: (b) x^* is the unique point of accummulation of $\{x^k\}$, and

(c) with respect to the grouping (20),

$$\lim_{c\to\infty} \frac{\|x_c^n - x^*\|}{\|x_c^0 - x^*\|} = 0, \tag{23}$$

i.e., convergence is superlinear with respect to every n points within a group.

Proof: Part (a) is just Theorem 1, part (b) follows from (8) and (22). The proof of part (c) will consist of a series of assertions.

There are two numbers α_1 and α_2 such that if λ^k is an eigenvalue of H^k, and k is large,

$$0 < \alpha_1 \leq \lambda^k, \tag{24}$$

also

$$\|H^k s\| \leq \alpha_2 \|s\|, \quad \text{for all } s. \tag{25}$$

This follows because the third term of (4) can be written $\sigma^k[(\sigma^k)^t \hat{G}(\eta^k)\sigma^k]^{-1}(\sigma^k)^t$ and in a neighborhood of x^*, the eigenvalues of $\hat{G}(\eta^k)$ are strictly positive and bounded below away from zero.

In a neighborhood about x^* there are two finite positive numbers α_6, α_7 such that for y in that neighborhood,

$$\alpha_6 \|y - x^*\| \leq \|g(y)\| \leq \alpha_7 \|y - x^*\|, \tag{26}$$

where $g(y) \equiv \nabla f(y)$. This is proved using a Taylor's expansion on $g(y)$, the stationarity of x^* and the positive definiteness of G^*.

In a neighborhood about x^*, there is a positive value α_8 such that for y, z in that neighborhood, if $f(y) \leq f(z)$, then

$$\|y - x^*\| < \alpha_8 \|z - x^*\| . \tag{27}$$

The proof follows like that in (26).

In a neighborhood about x^* there are values α_{11} and α_{12} such that,

$$0 < \alpha_{11} \leq t^k \leq \alpha_{12} . \tag{28}$$

This follows directly from (19), (24), (25), and the positive definiteness of G^*.

In a neighborhood about x^*, there are two values α_9, α_{10}, such that,

$$0 < \alpha_9 \leq \frac{\|s^k\|}{\|g^k\|} \leq \alpha_{10} . \tag{29}$$

This follows directly from (7), (24), and (25).

In a neighborhood of x^* there is a value α_{13} such that if y, z are in that neighborhood, and $f(y) \leq f(z)$, then

$$\|g(y)\| \leq \alpha_{13} \|g(z)\| . \tag{30}$$

In a neighborhood of x^* there are values α_{14}, α_{15} such that

$$0 < \alpha_{14} \leq \frac{\|y_c^k\|}{\|\sigma_c^k\|} \leq \alpha_{15} . \tag{31}$$

This is easily proved using (17) and the positive definiteness of G^*.

Let

$$\gamma = \limsup_{c \to \infty} \frac{\|x_c^n - x^*\|}{\|x_c^0 - x^*\|} \tag{32}$$

(Because of (27), $\gamma < +\infty$.) To prove part (c) we need to show that $\gamma = 0$. Let I^1 be an ordered subset of integers where

$$\gamma = \lim_{\substack{c \to \infty \\ c \in I^1}} \frac{\|x_c^n - x^*\|}{\|x_c^0 - x^*\|} .$$

Let

$$\delta = \liminf_{\substack{c \to \infty \\ c \in I^1}} \frac{\|g_c^{n-1}\|}{\|g_c^0\|} . \tag{33}$$

Let I^2 be an ordered subset (of I^1) such that

$$\delta = \liminf_{\substack{c \to \infty \\ c \in I^2}} \frac{\|g_c^{n-1}\|}{\|g_c^0\|} = \limsup_{\substack{c \to \infty \\ c \in I^2}} \frac{\|g_c^{n-1}\|}{\|g_c^0\|} .$$

There are two cases to consider.

(i) Suppose $\delta = 0$. Then because of (26), the fact that $f(x_c^n) < f(x_c^{n-1})$, and (27),

$$\lim_{\substack{c \to \infty \\ c \in I^2}} \frac{\|x_c^n - x^*\|}{\|x_c^0 - x^*\|} = 0 . \tag{34}$$

Hence $\gamma = 0$ for case (i).

(ii) Suppose $\delta > 0$. Here a modified induction proof is required. In the following it is assumed that only c's in I^2 are under consideration. For those $c \in I^2$, there is an $\alpha_{16} > 0$ such that

$$\|g_c^i\| \geq \alpha_{16} \|g_c^0\|, \quad 1 \leq i \leq n-1 . \tag{35}$$

In the following there are three propositions I, II, and III. In the modified induction proof (modified in the sense that the propositions are to be proved not for all the integers, just for a subset), superscripts (e.g. I^k) will be used to indicate the integer for which the truth of the proposition is being considered.

The three propositions are as follows.

I: $(g_c^k)' s_c^i = \|g_c^k\| \cdot \|s_c^i\| \cdot \epsilon_c^{k,i}$,

where

$$\lim_{c \to \infty} \epsilon_c^{k,i} = 0, \tag{36}$$

for $0 \leq i < k$; $k = 1, \ldots, n-1$.

For $k = n$,

$$(g_c^n)' s_c^i = \|g_c^0\| \cdot \|s_c^i\| \cdot \epsilon_c^{n,i}$$

where

$$\lim_{c \to \infty} \epsilon_c^{n,i} = 0 \tag{37}$$

for $i = 0, \ldots, n-1$.

II: $H_c^k G^* s_c^i = s_c^i + \delta_c^{k,i}$,

where

$$\lim_{c\to\infty} \frac{\|\delta_c^{k,i}\|}{\|s_c^i\|} = 0, \tag{38}$$

for $0 \le i < k$; $k = 1, \ldots, n-1$.

III: $(s_c^k)' G^* s_c^i = \|s_c^k\| \cdot \|s_c^i\| \cdot \gamma_c^{k,i}$

where

$$\lim_{c\to\infty} \gamma_c^{k,i} = 0, \tag{39}$$

for $0 \le i < k$; $k = 1, \ldots, n-1$.

(The relations implied by III are a generalization of those given by conjugate directions which hold when $f(x)$ is a positive definite quadratic form).

We note in passing that in arguments using limits as those stated above, if the conclusions are true for a finite collection of terms, they are true for their sum. This fact will be used implicitly in the following proof.

Proof for $k = 1$

I^1: $$(g_c^1)' s_c^0 = 0 \qquad \text{(from (14))}. \tag{40}$$

II^1: $$H_c^1 \hat{G}(\eta_c^0) s_c^0 = H_c^1 y_c^0 (t_c^0)^{-1} \qquad \text{(from (9), (17))}.$$

$$= s_c^0 \quad \text{(from (4), (9))}. \tag{41}$$

Thus

$$H_c^1 G^* s_c^0 = H_c^1 \hat{G}(\eta_c^0) s_c^0 + H_c^1 [G^* - \hat{G}(\eta_c^0)] s_c^0$$
$$= s_c^0 + H_c^1 [G^* - \hat{G}(\eta_c^0)] s_c^0 . \qquad (42)$$

Part II^1 follows from the fact that $[G^* - G(\eta_c^0)] \to 0$ as $c \to \infty$ and (25).

III^1: $(s_c^1)' G^* s_c^0 = (s_c^1)' \hat{G}(\eta_c^0) s_c^0 + (s_c^1)' [G^* - \hat{G}(\eta_c^0)] s_c^0$

$$= 0 \quad \text{(using (41) then (40))} \qquad (43)$$
$$+ (s_c^1)' [G^* - \hat{G}(\eta_c^0)] s_c^0 .$$

The remainder of the proof follows from the arguments used for part II^1.

<u>Assume true for k, prove true for k + 1</u>

I^{k+1}: A. Case where $2 \le k + 1 \le n - 1$.)

For $i = k$, $(g_c^{k+1})' s_c^k = 0$ (from (14)).

For $0 \le i \le k$,

$$(g_c^{k+1})' s_c^i = [g_c^k + \hat{G}(\eta_c^k) s_c^k t_c^k]' s_c^i \qquad \text{(Taylor's Theorem)}$$
$$= (g_c^k)' s_c^i + (s_c^k)' G^* s_c^i t_c^k + (s_c^k)' [\hat{G}(\eta_c^k) - G^*] s_c^i t_s^k . \qquad (44)$$

The induction hypothesis I^k, with (30) and (35) takes care of the first term in (44). The induction hypothesis III^k with (28), (30) and (35) takes care of the second term. Part I^{k+1} for the third term is trivial.

B. Case where $k + 1 = n$.

Using the fact that $f(x_c^{n-1}) < f(x_c^0)$ along with (30) and the induction hypotheses yields (37).

II^{k+1}: For $i = k$, the proof is identical to that of II^1. For $i < k$,

$$H_c^{k+1}G^*s_c^i = H_c^kG^*s_c^i - H_c^ky_c^k[(y_c^k)'H_c^ky_c^k]^{-1}(y_c^k)'H_c^kG^*s_c^i + \sigma_c^k[(\sigma_c^k)'y_c^k]^{-1}(\sigma_c^k)G^*s_c^i . \tag{45}$$

The induction hypothesis $II^k(H_c^kG^*s_c^i = s_c^i + \delta_c^{k,i})$ takes care of the first term in (45). Using it in the second term gives two more quantities, the second of which

$$-H_c^ky_c^k[(y_c^k)'H_c^ky_c^k]^{-1}(y_c^k)'\delta_c^{k,i}$$

yields the desired conclusion because the magnitude of y_c^k is an independent quantity, (24), (25), and the induction assumption II^k. For the first quantity

$$-H_c^ky_c^k[(y_c^k)'H_c^ky_c^k]^{-1}(y_c^k)'s_c^i ,$$

the desired conclusion will follow if

$$(y_c^k)'s_c^i = \|y_i^k\| \cdot \|s_c^i\| \cdot \mu_c^{k,i} , \tag{46}$$

where

$$\lim_{c\to\infty} \mu_c^{k,i} = 0. \tag{47}$$

Now

$$(y_c^k)'s_c^i = t_c^k(s_c^k)'\hat{G}(\eta_c^k)s_c^i \qquad \text{[Taylor's Theorem]}$$
$$= t_c^k(s_c^k)'G^*s_c^i + t_c^k(s_c^k)'[\hat{G}(\eta_c^k) - G^*]s_c^i \ . \tag{48}$$

Because of III^k, and the continuity of $G(x)$, both terms have the requirements necessary to show (46) and (47) hold. The conclusions for the third term in (45) follows similar arguments. This completes the proof of II^{k+1}.

III^{k+1}: For $i = k$,

$$(s_c^{k+1})'G^*s_c^k = (s_c^{k+1})'\hat{G}(\eta_c^k)s_c^k + (s_c^{k+1})'[G^* - \hat{G}(\eta_c^k)]s_c^k$$
$$= 0 + (s_c^{k+1})'[G^* - \hat{G}(\eta_c^k)]s_c^k \ . \tag{49}$$

The usual arguments apply.

For $i < k$,

$$(s_c^{k+1})'G^*s_c^i = (g_c^{k+1})'H_c^{k+1}G^*s_c^i \qquad \text{(by definition)}$$
$$= (g_c^{k+1})'(s_c^i + \delta_c^{k+1,i}) \qquad (\text{by } II^{k+1}) \ ,$$

which, using I^{k+1}, and the property on $\delta_c^{k+1,i}$ completes the proof of III^{k+1}.

We shall now show how the desired conclusions follow from I^k (eq. (37)).

First, note that because of III the matrix

$$S^c \equiv \{s_c^0 \|s_c^0\|^{-1}, \ldots, s_c^{n-1} \|s_c^{n-1}\|^{-1}\} \tag{50}$$

for all c large has an inverse, and

$$\liminf_{c \to \infty} |\det(S^c)| > 0.$$

This follows because it is easy to show from III that any limit set of vectors in the matrix S_c above must be linearly independent.

Then, using (37) and solving for g_c^n,

$$(g_c^n)' = \|g_c^0\|[\epsilon_c^{n,0}, \ldots, \epsilon_c^{n,n-1}](S^c)^{-1}. \tag{51}$$

Since the $\epsilon^{n,i}$'s all have property (37), for case (ii), part (c) of the theorem follows from (26) by taking the norm of both sides of (51). Q.E.D.

An obvious corollary of this theorem follows from II.

Corollary 1 [Convergence to the Inverse Hessian]

Under the assumptions (1), (2), and (3) of Theorem 2, for $c \in I^2$,

$$\lim_{c \to \infty} \|[H_c^n - (G^*)^{-1}]s\| = 0, \quad \text{for all } s. \tag{52}$$

If a Lipschitz condition is placed on the second derivatives of f it is possible to show that the rate of convergence to the strict local minimum is at least quadratic for certain subsets of integers.

Theorem 3. [Quadratic Rate of Convergence]

If: (1) in a neighborhood of x^* there is an α_3 such that for any y, z in that neighborhood, for any j, $(j = 1, \ldots, n)$

$$\left|\sum_{i=1}^{n} \left(\partial^2 f(y)/\partial x_i \partial x_j - \partial^2 f(z)/\partial x_i \partial x_j\right) s_i\right| \leq \frac{\alpha_3}{n} \cdot \|s\| \cdot \|y - z\|, \quad \text{[Lipschitz Condition on second derivatives of f]} \tag{53}$$

(2) the RDVMM is applied to problem (1),

(3) G^* is a positive definite matrix,

then: (a) There is an α_4 such that for c large,

$$\|x_c^n - x^*\| \leq \alpha_4 \|x_c^0 - x^*\|^2 \tag{54}$$

for $c \in I^3$, any ordered set of integers with the property that

$$\liminf_{\substack{c \to \infty \\ c \in I^3}} \frac{\|g_c^{n-1}\|}{\|g_c^0\|} = \delta > 0 . \tag{55}$$

(The qualification on the set of integers for which "at least quadratic" convergence can be proved is needed because if the gradient of f "drops an order of magnitude" during a group of iterations before the n th point, the induction step I^{k+1} (see equation (44)) fails. This drop in magnitude contributes to the superlinear convergence (see Theorem 2), but it may not be as high as quadratic.)

Proof: The proof of this theorem uses many results of the proof of Theorem 2. To avoid duplication, results from that theorem will be used. First, we note that case (ii) obtains since I^3 is a set of integers being the same properties as I^2 (compare(33) and (55)).

As in Theorem 2, three propositions need to be proved.

I: $(g_c^k)' s_c^i = \|g_c^k\| \cdot \|s_c^i\| \epsilon_c^{k,i}$, where

$$|\epsilon_c^{k,i}| \leq \alpha_c^{k,i} \|x_c^0 - x^*\| , \tag{56}$$

$$\limsup_{c \to \infty} \alpha_c^{k,i} < +\infty . \tag{57}$$

for $0 \leq i < k$; $k = 1, \ldots, n-1$.

For $k = n$,

$$(g_c^n)' s_c^i = \|g_c^0\| \cdot \|s_c^i\| \epsilon_c^{n,i} , \tag{58}$$

where $\epsilon_c^{n,i}$ have the properties implied by (56) and (57).

II: $H_c^k G^* s_c^i = s_c^i + \delta_c^{k,i}$,

where

$$\frac{\|\delta_c^{k,i}\|}{\|s_c^i\|} \leq \beta_c^{k,i} \|x_c^0 - x^*\|$$

where

$$\limsup_{c \to \infty} \beta_c^{k,i} < +\infty .$$

for $0 \leq i < k$; $k = 1, \ldots, n-1$.

III: $(s_c^k)' G^* s_c^i = \|s_c^k\| \cdot \|s_c^i\| \cdot \gamma_c^{k,i}$

where

$$|\gamma_c^{k,i}| \leq \omega_c^{k,i} \|x_c^0 - x^*\|$$

where

$$\limsup_{c \to \infty} \omega_c^{k,i} < +\infty,$$

for $0 \leq i < k$; $k = 1, \ldots, n-1$.

Analysis of these statements and a comparison with those of Theorem 2 show that the difference is that the terms that vanish in Theorem 2 are said to vanish (roughly) at the rate that $\|x_c^0 - x^*\| \to 0$.

To write out the complete proof would duplicate most of the proof of Theorem 2. It shall be sufficient to analyze that vanishing quantity of equations (44) which is not involved with the induction hypothesis. Thus

$$|(s_c^k)' [\hat{G}(\eta_c^k) - G^*] s_c^i t_c^k|$$

$$\leq \|s_c^k\| \cdot \|[\hat{G}(\eta_c^k) - G^*] s_c^i\| \cdot t_c^k \qquad (59)$$

$$\leq \|s_c^k\| \cdot \|s_c^i\| \frac{\alpha_3}{n} [\sum_{j=1}^{n} \|\eta_c^{k,j} - x^*\|] t_c^k . \qquad (59)$$

(using (15), (16), and (53)).

For $j = 1, \ldots, n$ it follows from (18) and (27) that

$$\|\eta_c^{k,j} - x^*\| \leq \alpha_8 \|x_c^k - x^*\| .$$

Using this in (59) the chain of inequalities continues as

$$\begin{aligned} &\leq \|s_c^k\| \cdot \|s_c^1\| \cdot \alpha_3 \cdot \alpha_8 \cdot \|x_c^k - x^*\| \cdot t_c^k \\ &\leq \|s_c^k\| \cdot \|s_c^1\| \cdot \alpha_3 \cdot \alpha_8^2 \cdot \alpha_{12} \|x_c^0 - x^*\| \end{aligned} \tag{60}$$

(using the fact that $f(x_c^k) < f(x_c^0)$ with (27), and (28)).

Thus (60) is of the form required in (56) and (57). Q.E.D.

Corollary 2.

Under the assumptions of Theorem 3, there is an α_{18} such that for c large, and $c \in I^3$,

$$\|[H_c^n - (G^*)^{-1}]z\| \leq \alpha_{18} \cdot \|x_c^0 - x^*\| \cdot \|z\|$$

for all z.

Proof: The proof is similar to that of Corollary 1.

The important observation about all this is that Theorems 2 and 3 on the rate of convergence would also apply if the resetting occurred at the nth and not the (n+1)th point. That is, if (10) were replaced with "$(k+1) \equiv 0 \bmod(n)$" instead of "$(k+1) \equiv 0 \bmod(n+1)$." This emphasizes the tentative conclusion of McCormick and Pearson [3] that the rate of convergence

of Davidon's Variable Metric Method depends on its conjugate direction properties not on the fact that if corollary 1 holds it is also a quasi-Newton method.

For the original DVMM Powell [5] has shown that convergence to a stationary point is guaranteed when the function to be minimized has a Hessian Matrix whose eigenvalues are bounded below away from zero. In Theorem 1 it was shown that the RVMM converges when just the continuity of the first derivatives is required. There is experimental evidence in McCormick and Pearson [3], to indicate that without the resetting feature, the DVMM can fail to converge for a nonconvex function.

In [5] Powell showed that the rate of convergence of the DVMM is every step superlinear if the second derivatives of f are Lipschitzian. Under the same assumption in Theorem 3 it was shown that the RVMM could be expected to exhibit a quadratic rate of convergence every n steps. Furthermore, with just the assumption that the eigenvalues of ∇f be bounded below away from zero, the RVMM has n-step superlinear convergence. In the first case it seems reasonable that an every step superlinear rate of convergence would be better than n-step quadratic rate. There is currently no theoretical analysis of this statement.

Acknowledgements

The author wishes to thank A. V. Fiacco, A. P. Jones, and P. D. Robers for their helpful criticisms and suggestions.

REFERENCES

1. Davidon, W. C., "Variable Metric Method for Minimization," Research and Development Report ANL-5990 (Revised), Argonne National Laboratory, U. S. Atomic Energy Commission, 1959.

2. Fletcher, R., and M. J. D. Powell, "A Rapidly Convergent Descent Method for Minimization," Computer Journal, 6: 163-16 , 1963.

3. McCormick, G. P., and J. D. Pearson, "Variable Metric Methods and Unconstrained Optimization," Proceedings of the Joint Conference on Optimization, University of Keele, March, 1968.

4. McCormick, G. P., "Anti-Zig-Zagging by Bending," Management Science, Vol. 15, No. 5, January 1969.

5. Powell, M. J. D., "On the Convergence of the Variable Metric Algorithm," T. P. 382, Theoretical Physics Division, A. E. R. E., Harwell, England, October, 1969.

AR 70-31 Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Mathematics Research Center, U. S. Army
University of Wisconsin, Madison, Wis. 53706

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
None

3. REPORT TITLE
THE RATE OF CONVERGENCE OF THE RESET DAVIDON VARIABLE METRIC METHOD

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Summary Report: no specific reporting period.

5. AUTHOR(S) (First name, middle initial, last name)
Garth P. McCormick

6. REPORT DATE
October, 1969

7a. TOTAL NO. OF PAGES
21

7b. NO. OF REFS
5

8a. CONTRACT OR GRANT NO.
Contract No. DA-31-124-ARO-D-462

b. PROJECT NO.
None

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
#1012

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
None

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES
None

12. SPONSORING MILITARY ACTIVITY
Army Research Office-Durham, N.C.

13. ABSTRACT

The rate of convergence of the Reset Davidon Variable Metric Method is shown to be always superlinear and sometimes quadratic in cycles of n points when the limit point of the minimizing sequence has a positive definite Hessian matrix.

DD FORM 1 NOV 68 1473

Unclassified
Security Classification

AGO 5698A