On the Convergence of a Class of Derivative-free Minimization Algorithms

CS75005-R

Ann-sheng Chien Ting[1]

Published April, 1975

1. Assistant professor, Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

On the convergence of a class of
derivative-free minimization algorithms

Ann-sheng Chien Ting[1]

Abstract

A convergence analysis is presented for a general class of derivative-free algorithms for minimizing a function $f(x)$ whose analytic form of the gradient and the Hessian is impractical to obtain. The class of algorithms accepts finite difference approximation to the gradient with step-sizes chosen according to the following rule: if x, $\bar{x}$ are two successive iterate points and h, $\bar{h}$ are the corresponding step-size, then the following two conditions are required:

$$(1) \quad \|\bar{h}\| \leq \min\left(C_1 \|\bar{x} - x\|^2, \|h\|\right) \quad \text{for some } 0 < C_1 < \infty$$
$$(2) \quad \|h\| - \|\bar{h}\| \leq C_2 \|\bar{x} - x\|^2 \quad \text{for some } 0 < C_2 < \infty$$

The algorithms also maintain an approximation to the second derivative matrix and require the change in x made by each iteration is subject to a bound that is also revised automatically. The convergence theorems have the features that the starting point x^1 needs not be close to the true solution and $f(x)$ needs not be convex. Furthermore, despite of the fact that the second derivative approximation may not converge to the true Hessian at the solution, the rate of convergence is still Q-superlinear. The theory is also shown to be applicable to a modification of Powell's dog-leg algorithm.

Keyword: derivative-free, minimization, Q-superlinear convergence.

1. Assistant professor, Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

I. Introduction

Recently, Powell [4] has proved some global and superlinear convergence properties on a class of algorithms for unconstrained minimization. The methods are iterative. Given a starting point x^1, they generate a sequence of points $\{x^k\}_{k=1,2,\ldots}$, which is intended to converge to the minimum point x^* of the objective function $f:R^n \to R$. The class of algorithms maintain an approximation to the second derivative matrix but they do require the first derivatives of $f(x)$ to be calculated at every iteration. Very often the first derivatives are either not available or else extremely expensive to evaluate. In such cases, the applicability of these methods will need to be reappraised.

In this paper, we construct a general class of derivative-free algorithms by modifying Powell's class of methods. The first derivatives are replaced by finite difference approximations. We also show that by properly choosing the approximation to the gradient, the class of derivative-free algorithms retains all the global and superlinearly convergence properties. The convergence theorems are proved to be applicable to a modification of Powell's dog-leg algorithm [3].

Section 2 has a complete description of the class of derivative-free algorithms considered in this paper. Then in section 3 we prove that global convergence properties held by each algorithm in the class and we prove that if the iterative point x^k tend to a limit at which the second derivative matrix G(x) of f(x) is positive-definite and G(x) is continuous in its neighborhood, then this point is a local minimum and the matrices $\{B_k\}_{k=1,2,..}$, the approximation to $\{G(x^k)\}_{k=1,\ 2...}$, are uniformly bounded even though the conditions on B_k are not very restrictive. Section 4 includes a superlinearly convergence theorem under the assumption that B_k may not converge to G(x) at the solution x*. Section 5 applies all the theorems to a modified dog-leg algorithm.

2. The class of derivative-free algorithms

Let $f: R^n \to R$ be the function we want to minimize. Suppose f is twice differentiable, and let g(x) be its gradient vector and G(x) be its second derivative matrix. In the algorithms, we use $\bar{g}(x,h)$, the finite-difference gradient defined by [1]. (For convenience, sometimes we use $\bar{g}^k$ to denote $\bar{g}(x^k,h^k)$.)

$$(\bar{g}(x,h))_i = \begin{cases} \dfrac{f(x+h^Te_i)-f(x)}{h^Te_i} & \text{if } h^Te_i \neq 0 \\ \dfrac{\partial f(x)}{\partial x_i} & \text{if } h^Te_i = 0 \end{cases} \quad (2.1)$$

in place of g(x), where $h \varepsilon R^n$ is the step-size which will satisfy two conditions given later. Given a starting point x^1, the algorithms will iteratively generate a sequence of points $x^k(k=2,3,4,...)$ which is intended to converge to the minimum of f(x), x*.

At the beginning of each iteration, a point x^k is available with a nxn symmetric matrix B_k, a step bound Δ^k and a step-size h^k. B_k is an approximation to $G(x^k)$ and Δ^k is an upper bound for the change of x^k at this iteration. Both are generated from the previous iteration except B_1 is any symmetric matrix and Δ^1 is any positive constant. Both will be revised at each iteration with some rules given later. The step-size h^k for the finite-difference gradient $\bar{g}(x^k,h^k)$ is chosen according to the following two conditions:

(i) $||h^k|| \leq \min(C_1||x^k-x^{k-1}||^2, ||h^{k-1}||)$ $\quad o<C_1<\infty$ (2.2)

(ii) $||h^{k-1}||-||h^k|| \leq C_2||x^k-x^{k-1}||^2$ $\quad o<C_2<\infty$ (2.3)

where h^{k-1} is the step-size of x^{k-1}, if $x^k \neq x^{k-1}$. Otherwise, let $h^k=h^{k-1}$.

The step-size h^1, corresponding to x^1, is chosen arbitrary. Algorithms will terminate when $g(x^k)$ is zero, and because we want to study the convergence properties as k increases, we can assume that $g(x^k)$ is never identically zero. It is proved in [1] that $g(x^k)\neq 0$ implies that there exists h^k such that $\bar{g}(x^k,h^k)\neq 0$. Hence in our algorithms, we further restrict the step-size h^k such that $\bar{g}(x^k,h^k)\neq 0$, for all k. Now, we can describe the procedures at the kth iteration of our algorithms step by step in the following way:

Step 1: Let the increment vector s^k be defined by

$$s^k = \begin{cases} B_k^{-1}\bar{g}(x^k,h^k) & \text{if } ||B_k^{-1}\bar{g}^k|| \le \Delta^k \text{ \& } B_k \text{ is positive-definite} \quad (2.4) \\ \text{any vector } s \text{ which satisfies } \phi(x^k+s)<f(x^k) & \text{otherwise} \quad (2.5) \end{cases}$$

where $\phi(x^k+s)$ is the quadratic approximation to $f(x^k+s)$:

$$\phi(x^k+s)=f(x^k) + s^T\bar{g}^k + \frac{1}{2}s^T B_k s \tag{2.6}$$

Furthermore, the increment s^k must satisfy the inequality

$$f(x^k)-\phi(x^k+s^k)\ge C_3||\bar{g}^k|| \min[||s^k||, \left|\left|\frac{\bar{g}^k}{B_k}\right|\right|] \tag{2.7}$$

where C_3 is a positive constant. Then we define x^{k+1} by

$$x^{k+1} = \begin{cases} x^k+s^k & \text{if } f(x^k+s^k)<f(x^k) \quad (2.8(a)) \\ x^k & \text{otherwise} \quad (2.8(b)) \end{cases}$$

Step 2: Check the convergence criterion:

$$||\bar{g}(x^k,h^k)||<\varepsilon \tag{2.9}$$

for some tolerance $\varepsilon>0$. If (2.9) is true, the algorithm stops. Otherwise go to Step 3.

Step 3: Prepare for the next iteration. Generate B_{k+1} from B_k by any rule which satisfies

$$||B_{k+1}||\le C_4+C_5 \sum_{i=1}^{k+1} ||s^i|| \tag{2.10}$$

where C_4, C_5 are positive constants.

If $f(x^k)-f(x^k+s^k)\ge C_6(f(x^k)-\phi(x^k+s^k))$ (2.11)

with $0<C_6<1$,

let Δ^{k+1} be any constant which satisfies

$$||s^k||\le\Delta^{k+1}\le C_7||s^k|| \tag{2.12}$$

where $C_7\ge1$.

If (2.11) fails, Δ^{k+1} will satisfy

$$C_8||s^k||\le\Delta^{k+1}\le C_9||s^k|| \tag{2.13}$$

where $0<C_8\le C_9\le1$

Moreover, we impose a fixed upper bound $\bar{\Delta}$ for Δ^k. Let h^{k+1} be any constant which satisfies (2.2), (2.3) and $\bar{g}(x^{k+1}, h^{k+1})\ne0$. Then go to the Step 1 of the (k+1)th iteration.

By our assumption that $\bar{g}(x^k,h^k)\neq 0$, we know there exists s^k such that (2.5) is true. Condition (2.7) is stronger than either (2.4) or (2.5). There is no problem when B_k is not positive-definite, since the s^k which minimize (2.6) will satisfy (2.7). However, it is proved in Section 5 that equations (2.4) and (2.7) are consistent too. The consistency of conditions (2.2)(2.3) is proved in [1]. Also, it is proved in [1] that there exists $h^{k+1}\varepsilon R^n$ which satisfies both (2.2)(2.3) and $\bar{g}(x^{k+1},h^{k+1})\neq 0$. Hence, every step of the kth iteration is well-defined. And all the derivative-free algorithms analyzed in the paper will proceed iteratively according to the above description. From Section 2 through Section 4, whenever we mention a sequence $\{x^k\}$, we mean the iterative sequence $\{x^k\}$ generated by any algorithm considered here. Without losing generality, we assume here and throughout this paper the vector norms are Euclidean, the matrix norms are subordinate to the vector norms. In an attempt to increase the readbility of the material we have used notations allowing intermediate results occurring in one proof to be used in subsequent proofs. For example, if k_6 is chosen greater than or equal to k_5 in a proof, k_5 may have been chosen in a previous proof.

3. Global Convergence of the Algorithms

First we want to prove under reasonable condition on f(x), each algorithm of the class provides the limit

$$\liminf ||\bar{g}(x^k,h^k)||=0 \tag{3.1}$$

no matter where starting point x^1 is.

<u>Theorem 3.1</u>: Suppose f(x) is bounded below and differentiable, g(x) is uniformly continuous on a convex hull of the level set $L(x^1)$ of the starting point x^1. Then the vectors $\bar{g}(x^k,h^k)$ (k = 1, 2, 3, . . .) are not bounded away from zero.

<u>Proof</u>: Although most of the proof follows the proof of Theorem 1 in [ ], for the sake of completeness we will not omit any part of it.

Let Σ' denote the sum over the iterations for which condition (2.11) is satisfied. Suppose (2.11) holds for k = p and fails for k = p + 1, ..., q then expressions (2.12), (2.13) and the fact that $||s^k||\le\Delta^k$ imply the bound

$$\sum_{i=p}^{q} ||s^i||\le||s^p||[1+C_7+C_7\ C_9+C_7\ C_9{}^2+\ .\ .\ .$$
$$+\ C_7\ C_9{}^{q-p-1}]\ \le\ ||s^p||\ [1+\frac{C_7}{1-C_9}].$$

Therefore, the following inequality

$$\sum_{i=1}^{k} ||s^i|| \le [1 + \frac{C_7}{1 - C_9}] \ [||s^1|| + \sum_{i=2}^{k}{}' ||s^i||] \tag{3.2}$$

holds. Thus we deduce from inequality (2.10) that there exists constants $C_{10}>0$ and $C_{11}>0$ such that $\{B_k\}$ satisfy the condition

$$||B_k|| \le C_{10} + C_{11} \sum_{k=1}^{k}{}' ||s^i|| \tag{3.3}$$

The fact that $f(x)$ is bounded below and (2.8) implies $f(x^{k+1}) < f(x^k)$ show that sum $\sum_{k=1}^{\infty}{}' [f(x^k) - f(x^{k+1})]$ is convergent. Because Σ' denotes the sum over iterations for which condition (2.11) is satisfied, the sum $\sum_{k=1}^{\infty}{}' (f(x^k) - \phi(x^k+s^k))$ is convergent. Thus by applying the elementary inequality

$$\min[|a|, \ |b|] \ge \frac{|ab|}{|a| + |b|} \tag{3.4}$$

we deduce from expression (2.7) that the sum

$$\sum_{k=1}^{\infty}{}' \frac{||s^k|| \ ||\bar{g}^k||^2}{||\bar{g}^k|| + ||s^k|| \ ||B_k||} \tag{3.5}$$

is also convergent. The theorem is proved by obtaining a contridiction if $\bar{g}^k$ satisfies the bound $||\bar{g}^k|| \ge C_{12}$ where $C_{12} > 0$. In this case (3.5) and $\Delta^k \le \bar{\Delta}$ imply that

$$\sum_{k=1}^{\infty}{}' \frac{||s^k||}{1 + (\bar{\Delta}/C_{12})(C_{10} + C_{11} \sum_{i=1}^{k}{}' ||s^i||)}$$

is finite. It follows from the fact if $\sum_{k=1}^{\infty}{}' \frac{a_k}{\sum_{i=1}^{k}{}' a_i}$ is finite then $\sum_{k=1}^{\infty}{}' a_k$ is finite. Thus (3.3) shows that $\sum_{k=1}^{\infty}{}' ||s^k||$ is finite then there exists a constant $C_{13} > 0$ such that $||B_k|| \le C_{13}$ for all k.

Moreover, from (3.2) we find the limit $||s^k|| \to 0$. Let k_1 be so large that for all $k \ge k_1$, we have $||s^k|| \le C_{12}/C_{13} \le \frac{||\bar{g}^k||}{||B_k||}$. Hence it follows from (2.7) that

$$f(x^k) - \phi(x^k + s^k) \ge C_3 \, ||\bar{g}^k|| \, ||s^k|| \tag{3.6}$$

for all $k \ge k_1$. Thus the equality (2.6) gives

$$\left| 1 + \frac{{s^k}^T \bar{g}^k}{f(x^k) - \phi(x^k+s^k)} \right| = \left| \frac{\frac{1}{2} {s^k}^T B_k s^k}{f(x^k) - \phi(x^k+s^k)} \right| \le \frac{\frac{1}{2} |{s^k}^T B_k s^k|}{C_3 ||\bar{g}^k|| \, ||s^k||}$$

Since B_k is uniformly bounded, $||\bar{g}^k||$ is bounded away from zero and s^k tends to zero, the right hand side tends to zero, hence

$$\lim_{k \to \infty} \frac{{s^k}^T \bar{g}^k}{\phi(x^k+s^k) - f(x^k)} = 1 \tag{3.7}$$

Let $k_2 \ge k_1$ such that for all $k \ge k_2$ the left hand side of (3.7) is at least 1/2. By (3.6) and (3.7) we have

$$-{s^k}^T \bar{g}^k \ge \frac{1}{2} (f(x^k) - \phi(x^k+s^k)) \ge \frac{C3}{2} ||\bar{g}^k|| \, ||s^k|| \tag{3.8}$$

for all $k \ge k_2$. Since, for all $i = 1, \ldots, n$

$$|\bar{g}_i(x^k,h^k) - g_i(x^k)| = |g_i(x^k + \theta_i h^{k^T} e_i) - g_i(x^k)|$$

$$\leq ||g(x^k + \theta_i h^{k^T} e_i) - g(x^k)||$$

where

$$0 \leq \theta_i \leq 1,$$

we have the following inequality:

$$|f(x^k + s^k) - f(x^k) - s^{k^T} \bar{g}^k| \leq | \int_{\theta=0}^{1} s^{k^T} (g(x^k + \theta s^k)$$

$$-g(x^k))\, d\,\theta| + ||s^k||\ ||\bar{g}^k - g^k||$$

$$\leq ||\ s^k\ ||\ \{\omega\ (||s^k||) + \sqrt{n} \sup_{1\leq i\leq n}$$

$$\{||g(x^k + \theta_i h^{k^T} e_i) - g(x^k)||\}\}$$

$$\leq ||s^k||\ (\omega\ (||s^k||) + \sqrt{n}\ \omega\ (||h^k||)).$$

Here, $w(\cdot)$ is the modulus of continuity of $g(x)$ which is finite by the fact that $g(x)$ is uniformly continuous. Thus, since

$||s^k|| \rightarrow 0$ and $||h^k|| \leq c_1\ ||s^{k-1}||^2$,

$$\lim_{k\to\infty} \frac{f(x^k) - f(x^k+s^k)}{||s^k||} = \lim_{k\to\infty} - \frac{s^{k^T} \bar{g}^k}{||s^k||} \qquad (3.9)$$

Since (3.8) and $||\bar{g}^k||$ is bounded away from zero show that this right hand side is bounded away from zero, the ratio of the left hand side to the right hand side of equation (3.9) tends to 1. Therefore equation (3.7) gives the limit

$$\lim_{k\to\infty} \frac{f(x^k) - f(x^k + s^k)}{f(x^k) - \phi(x^k + s^k)} = 1 \tag{3.10}$$

showing that the test (2.11) holds for all sufficiently large k. Thus (2.12) implies that $\Delta^{k+1} \geq ||s^k||$ for $k \geq k_3 > 0$. Since $||s^k||$ is either Δ^k or $||B_k^{-1}\bar{g}^k||$, and $||B_k^{-1}\bar{g}^k|| \geq C_{12}/C_{13}$ with the fact that $||s^k|| \to 0$, there exists $k_4 > 0$ such that

$$||s^k|| = \Delta^k \text{ for all } k \geq k_4$$

Hence, $||s^{k+1}|| = \Delta^{k+1} \geq ||s^k||$ for $k \geq \max(k_4, k_3)$. In other words, after a finite number of iterations, $||s^k||$ stops decreasing. Since $||s^k||$ is always positive, we cannot obtain $||s^k|| \to 0$. This is a contradiction. Therefore $||\bar{g}^k||$ cannot be bounded away from zero. Proof is completed.

From the above theorem, we know there is no need for x^1 to be close to the solution x*. The sequence $\{x^k\}$ will converge to x*, if one of the points x^k falls into a region where f(x) is locally convex and contain a local minimum and if the definition of x^{k+1} will keep the later points of the sequence $\{x^k\}$ in this region. Hence we have the following theorem:

Theorem 3.2: Let the hypotheses of Theorem 3.1 hold, and assume

(3) $f(x)$ is strictly convex in a closed neighborhood S of the local minimum x^*,

(4) there exists an integer $\sigma > 0$ such that for all $k \geq \sigma$, the iterate points x^k all lie in S. Then $\{x^k\}$ converges to x^*.

Proof: Let $\rho_1 = \inf ||x^k - x^*||$, $k \geq \sigma$. If $\rho_1 > 0$, then we define $\rho_2 > 0$ so large that $||x - x^*|| < \rho_2$ for all $x \in S$. Hence for all $k \geq \sigma_1 \geq \sigma$ $\rho_1 \leq ||x^k - x^*|| \leq \rho_2$. Set $\tilde{S} = \{x : \rho_1 \leq ||x - x^*|| \leq \rho_2\}$ and $\bar{f} = \min_{x \in \tilde{S}} f(x)$.

Since f is strictly convex on S and $\rho_1 > 0$, we have $\bar{f} > f(x^*)$ and

$$f(x^*) \geq f(x^k) + (x^k - x^*)^T g^k$$

$$\geq f(x^k) - ||x^k - x^*|| \, ||g^k||$$

$$\geq \bar{f} - \rho_2 \, ||g^k||$$

Hence, we deduce the bound

$$||g^k|| \geq (\bar{f} - f(x^*))/\rho_2 \qquad k \geq \sigma_1$$

By the way we choose h^k, we have

$$||\bar{g} \quad (x^k, h^k)|| \geq 0 \text{ for all } k \geq \sigma_1$$

Then we have a contridiction to Theorem 3.1. That implies $\inf ||x^k - x^*||$ $= 0$ for $k \geq \sigma$. Because $f(x)$ is continuous and $f(x^k)$ decreases monotonically, we deduce the limit

$$\lim_{k\to\infty} f(x^k) = f(x^*) \tag{3.11}$$

Now we want to prove that for any $\varepsilon > 0$ there exists $\sigma_3(\varepsilon)$ such that for all $k \geq \sigma_3(\varepsilon)$, $||x^k - x^*|| < \varepsilon$. Let us define $\hat{f}(\varepsilon) = \min\limits_{\substack{||x-x^*||>\varepsilon \\ x\varepsilon S}} f(x)$, ε is any positive number, then $\hat{f}(\varepsilon) > f(x^*)$ and (3.11) implies that there exists $\sigma_3(\varepsilon) > 0$ such that for all $k \geq \sigma_3(\varepsilon)$, $f(x^k)<\hat{f}(\varepsilon)$. Hence for all $k \geq \sigma_3(\varepsilon)$, $||x^k - x^*|| < \varepsilon$ since $x^k \varepsilon S$. This concludes the proof.

Note that Theorem 3.1 states that the algorithms will terminate because, for some k, $||\bar{g}^k|| \leq \varepsilon$, where ε is the tolerance. It does not claim that the sequence $x^k (k=1,2,3...)$ would converge if ε were set to zero. However, Theorem 3.2 tells us that it usually happen that condition $f(x^{k+1})<f(x^k)$ tends to prevent divergence from a local minimum, so it is common for the sequence to tend to a limit.

The sequence $\{B_k\}$ of the approximations to the second derivative matrices $\{G(x^k)\}$ is generated by a rule which satisfies a very loose condition (2.10). But we can prove they are uniformly bounded if the sequence x^k tend to a limit x^* where $G(x^*)$ is positive definite. First, we need the following two lemmas.

Lemma 3.3: Assume

(1) the sequence x^k converges to a limit point x^*,

(2) the Hessian $G(x)$ of $f(x)$ exists and is continuous in a neighborhood N_0 of x^*, $G(x^*)$ is positive-definite, then there exists an interger $k_5 > 0$ and positive constants m, M, C_{14} such that for all $k \geq k_5$

(i) $m \, ||y||^2 \leq y^T G(x^k) \, y \leq M \, ||y||^2$ for $y \in R^n$,

(ii) $m \, ||x^k - x^*|| \leq ||g(x^k)|| \leq M \, ||x^k - x^*||$,

(iii) $m/2 \, ||x^k - x^*||^2 \leq f(x^k) - f(x^*) \leq M/2 \, ||x^k - x^*||^2$,

(iv) $\dfrac{||g(x^k)||}{||s^k||} \geq C_{14}$ if $x^{k+1} \neq x^k$.

Proof: Since $G(x^*)$ is positive definite and $G(x)$ is continuous in a neighborhood N_0 of x^*, we can find another neighborhood N_1 of x^* such that for all $x \varepsilon N_1$, $G(x)$ is positive definite. Let $M \geq ||G(x)||$ for all $x \varepsilon N_1$ and $\tilde{m}$ be a lower bound for the eigenvalues of $G(x)$ for all $x \varepsilon N_1$, then we have

$$\tilde{m}\ ||y||^2 \leq \ y^T G(x)\ y\ \leq M\ ||y||^2 \text{ for all } x \varepsilon N_1.$$

Let $\bar{m}$ be an upper bound for $||G(x)^{-1}||$ for all $x \varepsilon N_1$ and $\varepsilon > 0$ be so small that for all $||x - x^*|| < \varepsilon$ we have

$$||G(x) - G(x^*)|| < \frac{1}{2\bar{m}}$$

Since x^k converges to x^* there exists an integer $k_5 > 0$ such that for all $k \geq k_5$, $x^k \varepsilon N_1$ and $||x^k - x^*|| < \varepsilon$. It follows from a well-known result that

$$||g(x)|| \leq \sup_{t\varepsilon[0,1]} ||G(tx + (1 - t)\ x^*)(x - x^*)||$$

and

$$||g(x) - g(x^*) - G(y)(x - x^*)|| \leq \sup\ ||\ G(tx + (1 - t)\ x^*)$$
$$- G(y)||\ ||x - x^*||$$

where $y = rx + (1 - r)\ x^*$ for any $r\varepsilon[0,1]$. Therefore, we have

$$||g(x^k)|| \leq M\ ||x^k - x^*|| \text{ for } k \geq k_5$$

and $||g(x^k)|| \geq \left(\frac{1}{||G(y)^{-1}||} - \sup_{t\varepsilon[0,1]} ||G(tx^k + (1 - t)\ x^k - G(y)||\right) ||x^k - x^*||$, because

$||tx^k - (1-t)x^* - (rx^k + (1-r)x^*)|| < \varepsilon$ for all $t \varepsilon (0,1)$, we have

$$\sup_{t\varepsilon(0,1)} ||G(tx^k + (1 - t)\, x^*) - G(y)|| < \frac{1}{2\bar{\bar{m}}}$$

Hence, by the fact that $\dfrac{1}{||G(y)^{-1}||} \geq \dfrac{1}{\bar{\bar{m}}}$

$$||g(x^k)|| \geq \frac{1}{2\bar{\bar{m}}} \;\; ||x^k - x^*||$$

for all $k \geq k_5$. Choose $m = \min(\frac{1}{2\bar{\bar{m}}}, \tilde{m})$, then we have proved (i) and (ii).

From the identity

$$f(x^k) - f(x^*) = \int_0^1 (1 - \theta)(x^k - x^*)^T \, G(x^* + \theta(x^k - x^*))$$

$$(x^k - x^*)\, d\theta$$

and inequality (i), we deduce the bound (iii). Finally, if $x^{k+1} \neq x^k$ by applying (2.8 (a)) and (iii), we obtain

$$\begin{aligned} ||s^k|| = ||x^{k+1} - x^k|| &\leq ||x^{k+1} - x^*|| + ||x^k - x^*|| \\ &\leq ||x^k - x^*|| + (2[f(x^{k+1}) - f(x^*)]/m)^{1/2} \\ &\leq ||x^k - x^*|| + (2[f(x^k) - f(x^*)]/m)^{1/2} \\ &\leq (1 + \sqrt{M/m}) \;\; ||x^k - x^*|| \end{aligned} \tag{3.12}$$

Thus it follows from (ii) that $||s^k|| \leq (1 + \sqrt{M/m}\,) \dfrac{||g^k||}{m}$. Let

$C_{14} = \dfrac{m}{1 + \sqrt{M/m}}$, then $\dfrac{||g^k||}{||s^k||} \geq C_{14}$ for all $k \geq k_5$ and $x^{k+1} \neq x^k$.

It follows directly from the above lemma that, if $\{x^k\}$ converges to x^* at which $G(x^*)$ is positive definite and $G(x)$ is continuous on a neighborhood around x^*, then x^* is a local minimum.

The following lemma provides a relation between the finite-difference derivative $\bar{g}(x,h)$ and the real derivative $g(x)$.

Lemma 3.4. Suppose f is twice differentiable in an open set $D\subset R^n$ g is the gradient of f which satisfies the Lipschitz condition on D, i.e.

$$||g(y) - g(x)|| \leq C_0 ||y-x|| \quad \text{with } C_0>0,$$

for all $y\epsilon D$, $x\epsilon D$ and G is the Hessian matrix of f. Then we have

$$||\bar{g}(x,h)-g(x)||\leq C_0||h|| \tag{3.13}$$

In particular (3.12) is true if $||G(x)||$ is bounded by C_0.

Proof: See [1].

Theorem 3.5: Let the assumptions of Lemma 3.3 hold. Then the sum $\Sigma||s^k||$ is convergent. Furthermore, $\{||B_k||\}$ is uniformly bounded.

Proof: Because $(f(x^k) - f(x^*))$ is a monotonically decreasing sequence, we obtain

$$\sum_{k=1}^{m} [(f(x^k) - f(x^*)) - (f(x^{k+1}) - f(x^*))]/ \sqrt{f(x^k) - f(x^*)}$$

$$\leq 2 \sum_{k=1}^{m} (\sqrt{f(x^k) - f(x^*)} - \sqrt{f(x^{k+1}) - f(x^*)})$$

$$<2(\sqrt{F(x^1) - f(x^*)} - \sqrt{f(x^{m+1}) - f(x^*)} < 2\sqrt{f(x^1) - f(x^*)}$$

Hence, the sum

$$\sum_{k=1}^{\infty} [f(x^k) - f(x^{k+1})] \,/\, \sqrt{f(x^k) - f(x^*)} \tag{3.14}$$

is convergent. Suppose $K' = \{k : k \geq k_5$ and x^{k+1} is defined by (2.8(a))$\}$

Then by applying Lemma 3.3(ii), (3.12), (3.13) with $C_0=M$ and the fact the step sizes $\{h^k\}_{k \in K'}$ satisfy (2.2) and (2.3), we get

$$\begin{aligned} ||\bar{g}(x^k, h^k)|| &\leq ||g(x^k)|| + M\,||h^k|| \\ &\leq M\,||x^k - x^*|| + M\,(||h^k|| - ||h^{k+1}|| + ||h^{k+1}||) \\ &\leq M\,||x^k - x^*|| + M\,(C_1 + C_2)\,||s^k||^2 \\ &\leq (M + M(C_1 + C_2)(1 + \sqrt{M/m})\,||s^k||)\,||x^k - x^*|| \\ &< \hat{M}\,||x^k - x^*|| \end{aligned} \tag{3.15}$$

where $\hat{M} > 0$. Choose ε_1 so small that

$$\hat{m} = m - M(C_1 + C_2)(1 + \sqrt{M/m})^2\varepsilon_1 > 0.$$

Then there exists $k_6 \geq k_5 > 0$ such that for all $k \geq k_6$, $||x^k - x^*|| < \varepsilon_1$.

Thus again with Lemma 3.3 (ii) and (3.12) we have

$$\begin{aligned} ||\bar{g}(x^k, h^k)|| &\geq ||g(x^k)|| - M\,||h^k|| \geq m\,||x^k - x^*|| \\ &\quad - M(||h^k|| - ||h^{k+1}|| + ||h^{k+1}||) \end{aligned}$$

$$\geq m \, ||x^k - x^*|| - M\,(C_1 + C_2)\,||s^k||^2$$

$$\geq (m - M(C_1 + C_2)(1 + \sqrt{M/m})^2 ||x^k - x^*||)\,||x^k - x^*||$$

$$\geq (m - M(C_1 + C_2)(1 + \sqrt{M/m})^2\,\varepsilon_1)\,||x^k - x^*||$$

Then $||\bar{g}(x^k, h^k)|| \geq \hat{m}\,||x^k - x^*||$ (3.16)

for $k \geq k_6$ and $k \in K'$. It follows from (3.16) and Lemma 3.3 (iii) that

$$\sqrt{f(x^k) - f(x^*)} \leq \sqrt{M/2\hat{m}^2}\quad ||\bar{g}(x^k, h^k)||$$

for $k \geq k_6$ and $k \in K'$. Therefore, the summation (3.14) is convergent

implies that

$$\sum_{k=1}^{\infty} \frac{f(x^k) - f(x^{k+1})}{||g^k||}$$

is convergent. With equation (2.11) and (2.7) we obtain

$$\Sigma' \min[||s^k||, ||\bar{g}^k||/||B_k||] < +\infty .$$

Thus by applying (3.4), we deduce the sum

$$\Sigma' \frac{||s^k||\;||\bar{g}^k||}{||\bar{g}^k|| + ||s^k||\;||B_k||} \tag{3.17}$$

is also convergent. Remember that it followed from (3.12) that

$$||s^k|| \leq (1 + \sqrt{M/m}\,)\,\frac{||\bar{g}^k||}{\hat{m}}$$

Hence, there is a constant $C_{15} = \dfrac{1 + \sqrt{M/m}}{\hat{m}} > 0$ such that

$$\left\| \frac{s^k}{g^k} \right\| \le C_{15}$$

for all $k \ge k_6$ and $k \in K'$. Since K' includes all the k in Σ', (3.3) and (3.17) show that

$$\infty > \Sigma' \frac{s^k}{1 + C_{15}\|B_k\|} \ge \Sigma' \frac{\|s^k\|}{1 + C_{10} + C_{11} \overset{k}{\Sigma'} \|s^i\|}$$

By the fact that if $\sum_{k=1}^{\infty} \frac{a_k}{\sum_{i=1}^{k} a_i}$ is finite then $\Sigma\, a_k$ is also finite, we have $\Sigma' \|s^k\|$ is finite. Therefore, because of inequality (3.2) $\Sigma \|s^k\|$ is finite. It follows directly from (2.10) that $\|B_k\|$ is uniformly bounded by $C_4 + C_5 \Sigma \|s^k\|$. The proof is completed.

4. Superlinear Convergence

Let us first state the following result which is proved in [1].

Theorem 4.1: Assume

(1) $g: R^n \to R^n$ be differentiable on an open convex set in R^n.

(2) for some x^* in D, g' is continuous at x^* and $g(x^*)$ is nonsingular.

(3) $\{B_k\}$ is a sequence of nxn nonsingular matrices,

(4) $\bar{g}(x,h)$ is an approximation rule for $g(x)$ and suppose for some $x^1 \in D$ the sequence $\{x^k\}$ where $x^k = x^{k-1} - B_{k-1}^{-1}\, \bar{g}(x^{k-1}, h^{k-1})$ remains in D and converges to x^*,

then $\{x^k\}$ converges Q- superlinearly to x* and g(x*) = 0 iff

$$\lim_{k\to\infty} \frac{||(B_k - G(x^*))(x^{k+1} - x^k) + \bar{g}(x^k, h^k) - g(x^k)||}{||x^{k-1} - x^k||}$$

Proof: See [1].

With this result, we can prove the superlinearly convergence property of our class of algorithms even if B_k may not converge to G(x*).

Theorem 4.2: Let the hypotheses of Lemma 3.3 hold and in addition assume

(4) the matrices B_k satisfy the following condition

$$\lim_{k\to\infty} \frac{||(B_k - G(x^*))\, s^k + \bar{g}(x^k, h^k) - g(x^k)||}{||s^k||} = 0 \qquad (4.1)$$

Then $\{x^k\}$ converges Q-superlinearly to x*.

Proof: The assumption (4) implies that there exists $k_7 \geq k_6$ such that for all $k \geq k_7$

$$||(B_k - G(x^*))\, s^k + \bar{g}(x^k, h^k) - g(x^k)|| < \frac{1}{2} m\, ||s^k|| \qquad (4.2)$$

where m is the constant in Lemma 3. 3. Therefore, by (2.5) (4.2) and Lemma 3.3 (i), we obtain

$$\begin{aligned} 0 &< f(x^k) - \phi\,(x^k + s^k) = -s^{k^T} \bar{g}^k - 1/2\, s^{k^T} B_k\, s^k \\ &\leq -s^{k^T} \bar{g}^k - 1/2\, s^{k^T} [G(x^*)\, s^k - (\bar{g}^k - g^k)] + \frac{m}{4} ||s^k||^2 \\ &\leq -s^{k^T} \bar{g}^k - 1/2\, s^{k^T} G(x^*)\, s^k + 1/2\, s^{k^T} [\bar{g}^k - g^k] + \frac{m}{4} ||s^k||^2 \end{aligned}$$

$$\leq - s^{k^T} \bar{g}^k - 1/4\ m\ ||s^k||^2 + 1/2\ ||s^k||\ ||\bar{g}^k - g^k|| \tag{4.3}$$

If $x^{k+1} \neq x^k$, then it follows from (3.12) and h^k satisfy (2.2)(2.3) that

$$||\bar{g}^k - g^k|| \leq M\ ||h^k|| \leq M\ (||h^k|| - ||h^{k+1}|| + ||h^{k+1}||)$$

$$\leq M\ (C_1 + C_2)\ ||s^k||^2 \tag{4.4}$$

However, since x^k is convergent to x^*, if $x^{k+1} = x^k$ is defined by (2.8(b)) we can always find a x^{k+i}, $i > 1$ such that $x^{k+i} \neq x^k$ and $x^{k+i-j} = x^k$ for all $1 \leq j \leq i - 1$. In this case we have $h^k = h^{k+1} = . . ., = h^{k+i-1}$ and

$$||h^{k+i-1}|| - ||h^{k+i}|| \leq C_2\ ||s^{k+i-1}||^2$$

$$||h^{k+i}\ || \leq C_1\ ||s^{k+i-1}||^2$$

Because $x^k = x^{k+i-j}$, $1 \leq j \leq i - 1$, the inequality (2.11) must be failed for x^{k+i-j} with $1 \leq j \leq i - 1$. Therefore, by (2.13) we obtain

$$||s^{k+i-1}|| \leq \Delta^{k+i-1} \leq C_9\ ||s^{k+i-2}|| \leq C_9 \Delta^{k+i-2} \leq . . . \leq c_9^{\ i-1} ||s^k||$$

Since $C_9 < 1$, the above inequalities imply that

$$||s^{k+i-1}|| \leq ||s^k||$$

which gives

$$||h^k|| - ||\ h^{k+i}|| \leq C_2 ||s^k||^2$$

$$||h^{k+i}|| \le C_1 \; ||s^k||^2$$

Hence, (4.4) can be proved even if s^k is not used to define x^{k+1}, i.e., $x^k = x^{k+1}$ is defined by (2.8(b)). Therefore, (4.3) gives the following inequality

$$0 < -s^{k^T} \bar{g}^k - 1/4 \; m \; ||s^k||^2 + 1/2 \; M(C_1 + C_2) \; ||s^k||^3.$$

If we further choose $k_8 \ge k_7$ so large that for all $k \ge k_1$

$$M \; (C_1 + C_2) \; ||s^k|| \le m/8,$$

then we have

$$0 < -s^{k^T} \; \bar{g}^k - 1/4 \; m \; ||s^k||^2 + \frac{1}{8} m \; ||s^k||^2 < -s^{k^T} \bar{g}^k - \frac{m}{8} \; ||s^k||^2$$

for all $k \ge k_8$, which gives the inequality

$$||\bar{g}^k|| \ge \frac{1}{8} m \; ||s^k|| \qquad\qquad k \ge k_8$$

Therefore, from (2.7) we obtain

$$f(x^k) - \phi(x^k + s^k) \ge C_3 \; \min[\frac{1}{8} m \; ||s^k||^2, \; \frac{(m \; ||s^k||)^2}{64 \; ||B_k||}].$$

It follows from (3.3) and Theorem 3.5 that there exists $C_{16} \ge 0$ such that $||B_k|| \le C_{16}$ for $k \ge k_4$. Hence, there exists a positive constants C_{17} such that

$$f(x^k) - \phi(x^k + s^k) \ge C_{17} \; ||s^k||^2 \qquad (4.5)$$

for $k \ge k_8$.

From Taylor's Theorem we can deduce

$$f(x^k + s^k) - f(x^k) = g^{k^T} s^k + \int_0^1 s^{k^T} G(x^k + ts^k) s^k (1 - t)\, dt.$$

Since $\int_0^1 s^{k^T} G(x^*) s^k (1 - t)\, dt = 1/2\, s^{k^T} G(x^*) s^k$ this provides

$$f(x^k + s^k) - f(x^k) = g^{k^T} s^k + \int_0^1 s^k [G(x^k + ts^k) - G(x^*)] s^k (1 - t)\, dt + \frac{1}{2} s^{k^T} G(x^*) s^k$$

From (2.5) we have

$$f(x^k) - \phi (x^k + s^k) = -\bar{g}^{k^T} s^k - \frac{1}{2} s^{k^T} B_k s^k$$

Let us assume that $f(x^k + s^k) - \phi(x^k + s^k)$ is positive. Add the above two identities, we obtain

$$f(x^k + s^k) - \phi(x^k + s^k) = (g^k - \bar{g}^k)^T s^k + \frac{1}{2} s^{k^T} (G(x^*) - B_k) s^k + \int_0^1 s^{k^T} [G(x^k + ts^k) - G(x^*)] s^k (1-t)\, dt$$

$$\leq \frac{1}{2} ||g^k - \bar{g}^k|| \, ||s^k|| + \frac{1}{2} || (B_k - G(x^*)) s^k + \bar{g}^k - g^k || \, ||s^k|| + \frac{1}{2} ||s^k||^2 \max_{t\varepsilon[0,1]} ||G(x^k + ts^k) - G(x^*)||$$

By (4.4),

$$0 < \frac{f(x^k + s^k) - \phi (x^k + s^k)}{||s^k||^2} \leq M(C_1 + C_2) ||s^k|| +$$

$$\frac{1}{2} \quad \frac{||(B_k - G(x^*))\ s^k + \bar{g}^k - g^k||}{||s^k||} \quad +$$

$$\frac{1}{2} \max_{t \varepsilon [0,1]} \ ||G(x^k + ts^k) - G(x^*)||$$

As $||s^k|| \to 0$, $x^k \to x^*$ so $x^k + ts^k \to x^*$, by the continuity of $G(x)$ and

(4.1)

$$\frac{f(x^k + s^k) - \phi(x^k + s^k)}{||s^k||^2} \quad \to \quad 0 \tag{4.6}$$

Since

$$\frac{f(x^k) - \phi(x^k + s^k)}{f(x^k) - f(x^k + s^k)} \quad = \quad \frac{f(x^k) - f(x^k + s^k) + f(x^k + s^k) - \phi(x^k + s^k)}{f(x^k) - f(x^k + s^k)}$$

$$= 1 + \frac{(f(x^k + s^k) - \phi(x^k + s^k)/||s^k||^2}{\dfrac{f(x^k) - \phi\ (x^k + s^k)}{||s^k||^2} - \dfrac{f(x^k + s^k) - \phi(x^k + s^k)}{||s^k||^2}}$$

with (4.5) and (4.6) we have

$$\frac{f(x^k) - \phi(x^k + s^k)}{f(x^k) - f(x^k + s^k)} \to 1 \qquad \text{as } k \to +\infty$$

Thus there exists an $k_9 \geq k_8$ such that for all $k \geq k_9$ inequality (2.11) is satisfied if $f(x^k + s^k) > \phi(x^k + s^k)$. However, if $f(x^k + s^k) \leq \phi(x^k + s^k)$, (2.11) is by all means true. So for all $k \geq k_9$, (2.11) is satisfied, and the conditions

$$\begin{cases} \Delta^{k+1} \geq ||s^k|| \\ x^{k+1} = x^k + s^k \end{cases} \qquad \text{for all } k \geq k_9$$

hold. Therefore, if an iteration gives the reduction $||s^{k+1}|| \leq ||s^k||$, then the rule governing the definition of s^{k+1} implies that $s^{k+1} = -B_{k+1}^{-1}\bar{g}^{k+1}$. Since $||s^k||$ tends to zero as k tend to infinity, it follows that the Newton's formula is applied an infite many times.

Suppose, $k_{10} \geq k_9$ be so large that s^{k10} is defined by Newton's formula, and such that for $k \geq k_{10}$

$$||G(x^*)\ s^k + g(x^k) - g(x^k + s^k)|| < \frac{C_{14}\ ||s^k||}{2}$$

and

$$||(B_k - G(x^*))\ s^k + \bar{g}^k - g^k|| \leq \frac{C_{14}}{2}\ ||s^k||$$

(Note: C_{14} is defined in Lemma 3.3).

Add the above two inequalities and we have for all $k \geq k_{10}$

$$||B_k s^k + \bar{g}^k - g(x^k + s^k)|| < C_{14}\ ||s^k|| \tag{4.7}$$

Let $k \geq k_{10}$ be an integer such that $s^k = -B_k^{-1}\bar{g}^k$ then $x^{k+1} = x^k + s^k$ since $k \geq k_9$. Furthermore, (4.7) implies

that

$$||g(x^{k+1})|| < C_{14}\ ||s^k||$$

By Lemma 3.2(iv), we obtain

$$\frac{||g^{k+1}||}{||s^{k+1}||} > C_{14} > \frac{||g^{k+1}||}{||s^{k}||}$$

Hence $||s^{k+1}|| \leq ||s^{k}||$. It follows by induction that $s^k = -B_k^{-1}g^k$ for all $k \geq k_{10}$. Therefore, the subsequence $\{x_k\}_{k\geq k_{10}}$ satisfies all the assumptions of Theorem 4.1. This implies that the sequence $\{x_k\}_{k\geq k_{10}}$ converges to x* Q-superlinearly. Hence the sequence $\{x_k\}$ converges to x* Q-superlinearly.

5. Application

The above theorems prove all the convergence properties of the class of algorithms we discussed in Section 2. This theory can be applied to a practical method, a modification of Powell's dog-leg method [3]. (We call it algorithm MD).

Let us describe algorithm MD more precisely. In general algorithm MD will generate a sequence $\{x^k\}$ iteratively the way we defined in Section 2. However, the increment vector s^k, the step bounds $\{\Delta^k\}$ and the approximation matrices B_k to the Hessians are defined more specifically in the following way:

$$s^k = \begin{cases} -(B_k^{-1}\bar{g}(x^k, h^k)) \text{ if } B_k \text{ is positive definite and} \\ \qquad ||B_k^{-1}\bar{g}^k|| < \Delta^k, \\ -\Delta^k \dfrac{\bar{g}(x^k, h^k)}{||\bar{g}(x^k, h^k)||} \quad \text{if } (\bar{g}^{k^T} B_k \bar{g}^k)\, \Delta^k \le ||\bar{g}^k||^3, \\ (1-\alpha) \dfrac{||\bar{g}^k||^2\, \bar{g}^k}{\bar{g}^{k^T} B_k \bar{g}^k} + \alpha\, B_k^{-1}\bar{g}^k = s^k(\alpha), \text{ otherwise.} \\ \text{where } \alpha \text{ is choosen so that } \phi(x^k + s^k(\alpha)) \text{ is the least} \\ \text{for } ||s^k(\alpha)|| \le \Delta^k. \end{cases}$$

where $\bar{g}^k = \bar{g}(x^k, h^k)$, and h^k is chosen according to Section 2.

The step bound Δ^{k+1} is calculated according to the success of the kth iteration,

$$\Delta^{k+1} = \begin{cases} ||s^k|| \text{ or } 2||s^k|| \text{ if } f(x^k + s^k) - f(x^k) \le 0.1\ (\phi(x^k + s^k) - f(x^k)), \\ \frac{1}{2}\ ||s^k|| \quad \text{otherwise.} \end{cases}$$

And B_{k+1} is generated by:

$$B_{k+1} = B_k + \theta\, \frac{(y^k - B_k s^k)\, s^{k^T} + s_k (y^k - B_k s^k)^T}{||s^k||^2} - \theta^2\, \frac{s^{k^T}(y^k - B_k s^k)\, s^k s^{k^T}}{||s^k||^4} \tag{5.1}$$

where θ is the number closest to 1 such that $|\det B_{k+1}| > 0.1 |\det B_k|$,

$y^k = \bar{g}(x^k + s^k, h) - \bar{g}(x^k)$ where h is one of those h^{k+1}'s which satisfies (2.2) and (2.3) with $x^{k+1} = x^k + s^k$.

Now we will show that algorithm MD belongs to the class of derivative-free algorithms defined in Section 2 and hence also has all those desirable convergence properties proved in the previous sections. First we need the following theorem.

Theorem 5.1: Let $\eta \in R^n$ be the value of s that minimizes $\phi(x+s)$ subject to the inequality

$$||\eta^k|| \leq ||s^k|| \tag{5.2}$$

and subject to the condition that has the form

$$\eta^k = -\alpha\ \bar{g}^k \tag{5.3}$$

Then the bound

$$f(x^k) - \phi(x^k + \eta^k) \geq \frac{1}{2}\ ||\bar{g}^k||\ \min[||s^k||, \frac{||\bar{g}^k||}{||B_k||}] \tag{5.4}$$

is obtained

Proof: By (2.6) and (5.2) we have

$$f(x^k) - \phi(x^k + \eta^k) = \alpha\ ||\bar{g}^k||^2 - \frac{1}{2}\alpha^2\ \bar{g}^{k^T}\ B_k\ \bar{g}^k \tag{5.5}$$

Therefore, if $\bar{g}^{k^T}\ B_k\ \bar{g}^k \leq 0$, then the required vector η^k is obtained when α has the value $\alpha = \frac{s^k}{\bar{g}^k}$ in which case

$$f(x^k) - \phi\ (x^k + \eta^k) \geq ||s^k||\ ||\bar{g}^k||$$

which is consistent with the bound (5.4). However, if ${\bar{g}^k}^T B_k \bar{g}^k \geq 0$, then α is the number which will make the derivate of $(f(x^k) - \phi(x^k + \eta^k))$ w. r. t. α become 0, if $||\eta^k|| \leq ||s^k||$ still holds. Hence, $\alpha =$ $\min \left(\frac{||s^k||}{||\bar{g}^k||}, ||\bar{g}^k||^2/{\bar{g}^k}^T B_k \bar{g}^k\right)$. Thus,

$$\alpha {\bar{g}^k}^T B_k \bar{g}^k \leq ||\bar{g}^k||^2 \tag{5.6}$$

and by $\bar{g}^k B_k \bar{g}^k \leq ||\bar{g}^k||^2 ||B_k||$ we have,

$$\alpha \geq \min[||s^k||/||\bar{g}^k||, 1/||B_k||] \tag{5.7}$$

Thus, in this case, it follows from (5.5) (5.6) and (5.7) that

$$f(x^k) - \phi(x^k + \eta^k) \geq \frac{1}{2} \alpha ||\bar{g}^k||^2$$

$$\geq \frac{1}{2} ||\bar{g}^k|| \min[||s^k||, ||\bar{g}^k||/||B_k||]$$

The theorem is proved.

This theorem shows that the s^k defined by (2.4) will satisfy condition (2.7). Therefore, the s_k in Section 2 is well-defined. This theorem also proves that requiring condition (2.7) is equivalent to requiring the difference $f(x^k) - \phi(x^k + s^k)$ to be no less than a positive constant multiple of the greatest value of the difference $f(x^k) - \phi(x^k + \eta^k)$ that can be obtained when η^k is subject to (5.2) and (5.3). Algorithm MD defines the increment s^k in three ways. If s^k is defined either by

Newton's formula or as a gradient step, (2.7) is clearly satisfied. If s^k is defined by the combination of Newton's formula and the gradient step such that $\phi(x^k + s^k)$ will have the least value, then $\phi(x^k + s^k) \le \phi(x^k + \eta^k)$ for all η^k defined by (5.2), (5.3). Thus (2.7) is satisfied by this s^k too. Dennis [2] has proved that under the assumption $||G(x) - G(y)|| \le L\ ||x - y||$ the update of the Powell symmetric Broyden's method satisfies (2.10). Applying his proof by replacing $g(x)$ by $\bar{g}(x, h)$ we can prove that the matrices B_k of Algorithm MD satisfy (2.10). Hence, we are ready for the following global convergence theorem for Algorithm MD.

Theorem 5.2: Suppose $f(x)$ is bounded below and twice differentiable and $g(x)$ is uniformly continuous and there exists $L > 0$ such that

$$||G(x) - G(y)|| \le L\ ||x - y||$$

for all x, y in a convex hull of the level $L(x^1)$, where x^1 is the initial point. Then the vector sequences $\bar{g}(x^k, h^k)$ where $\{x^k\}$ is generated by Algorithm MD are not bounded away from the zero vector.

Theorem 5.3: Under the assumptions of Theorem 3.2, the sequence $\{x^k\}$ generated by Algorith MD will converge to the local minimum x^*.

Proof: Immediately follows from Theorem 3.2.

Furthermore, we can prove Algorithm MD actually converges Q-superlinearly.

Theorem 5.4: If the hypotheses of Theorem 5.2 hold and assume the sequence $\{x^k\}$ generated by Algorithm MD converges to x^*. Then $\{x^k\}$ is Q-superlinearly convergent to x^*.

Proof: Since the B_k's in Algorithm MD is generated by the Powell symmetric Broyden update, it is proved in [1] that B_k satisfies condition (4.1). Therefore, we can apply Theorem 4.2 to Algorithms MD. The result follows directly.

In the original dog-leg algorithm, every third iteration is a special iteration, for which s^k is defined in such a way that for some constant $E < 1$,

$$\left\| \prod_{j=k}^{k+\ell} \left(I - \theta^j \frac{s^i s^{j^T}}{\|s^j\|^2}\right) \right\|_2 \leq E$$

will be true for all k and a fixed ℓ. And it sets $\Delta^{k+1} = \Delta^k$. If we also adopt the special iterations into our Algorithm MD, (call it Algorithm MDS), then the convergence still can be retained.

Theorem 5.5: The result of Theorem 3.1 is true for Algorithm MDS, under the same assumptions.

Proof: The only difference between Algorithm MDS and the general class of algorithms is the special iterations. This proof will be concluded if we can prove that the special iterations do not affect the proof of Theorem 3.1. Let us redefine Σ' to denote the sum over the ordinary iterations for which (2.11) is satisfied. Because in a special iteration, $||s^k|| \le \Delta^k$ and $\Delta^{k+1} = \Delta^k$, the inequality (3.2) will still hold for the new definition of Σ'. Hence the rest of the proof of Theorem 3.1 applies unchanged to this theorem.

The following theorem shows the rate of convergence after the special iteration has been added.

Theorem 5.6: If the hypotheses of Theorem 5.4 hold and assume the sequence $\{x^k\}$ generated by Algorithm MD converges to x^*. Then $\{x^k\}$ is two-step Q-superlinearly convergent to x^*.

Proof: By Theorem 5.4 we know that x^ks generated only by ordinary iterations converge Q-superlinearly. The proof is completed if we can show the superlinear convergence is not damaged by the special iterations. It follows from Lemma 3.3(iii) that since $\{f(x^k) - f(x^*)\}$ is monotonically decreasing

$$||x^{k+1} - x^*|| \le \sqrt{M/m}\; ||x^k - x^*||$$

Because every special iteration is followed by two ordinary iterations, if kth iteration is a special iteration, (k-1)th, (k-2)th, (k+1)th, (k+2)th are ordinary iterations. Hence, the ratio

$$0 \le \frac{||x^k - x^*||}{||x^{k-2} - x^*||} \le \sqrt{M/m} \frac{||x^{k-1} - x^*||}{||x^{k-2} - x^*||}$$

tends to zero when k tends to infinity. And

$$0 \le \frac{||x^{k+2} - x^*||}{||x^k - x^*||} \le \frac{||x^{k+2} - x^*||}{\sqrt{m/M}\ ||x^{k+1} - x^*||}$$

tends to zero when k tends to infinity. Therefore, the ratio $\frac{||x^{k+2} - x^*||}{||x^k - x^*||}$ tends to zero as k tends to infinity for the sequence $\{x^k\}$ generated by Algorithm MDS. This implies that Algorithm MDS is two-step superlinearly convergent.

(1) Chien, Ann-sheng A., A class of derivative-free algorithms for unconstrained minimization, Cornell Computer Science Technical Report. TR74-220.

(2) Dennis, J. E., Jr., On some methods based on Broyden's secant approximation to the Hessian in Numerical methods for nonlinear optimization, edited by F.A. Loosma, Academic Press, 1972.

(3) Powell, M. J. D., A new algorithm for unconstrained optimization in Nonlinear programming, edited by J. B. Rosen, O. L. Mangasarian, K. Ritter, Academic Press, New York (1970)

(4) Powell, M. J. D., Convergence properties of a class of minimization algorithms, SIGMAP Nonlinear Programming Symposium, Univ. of Wisconsin, 1974.