ISSN: 0128-7680
Pertanika J. Sci. & Technol. 7(2): 111-123 (1999)

A Switching Criterion in Hybrid Quasi-Newton BFGS – Steepest Descent Direction

Malik Hj. Abu Hassan, Mansor B. Monsi and Leong Wah June
Department of Mathematics,
Faculty of Science and Environmental Studies,
Universiti Putra Malaysia,
43400 UPM , Serdang , Selangor Darul Ehsan, Malaysia

Received: 30 March 1998

ABSTRAK

Diperkenalkan dua kaedah terubahsuai untuk pengoptimuman tak berkekangan. Kaedah tersebut menggunakan strategi gabungan linear cembung dengan arah gelintaran tercuram quasi-Newton BFGS dan penurunan tercuram sebagai arah gelintaran. Suatu kriteria penukaran telah diterbitkan berdasarkan Syarat Peringkat Pertama dan Kedua Kuhn-Tucker. Kriteria penukaran tersebut boleh dilihat sebagai suatu cara pergerakan di antara langkah quasi-Newton dan penurunan tercuram berpadanan dengan syarat Kuhn-Tucker. Ini adalah untuk menjamin bahawa tiada gerakan tersaur berpotensi menjauhi langkah penurunan semasa yang mengurangkan nilai fungsi matlamat dapat dibuat. Keputusan berangka juga dipersembahkan dan menunjukkan bahawa kemajuan telah diperolehi berbanding dengan algoritma BFGS.

ABSTRACT

Two modified methods for unconstrained optimization are presented. The methods employ a hybrid descent direction strategy which uses a linear convex combination of quasi-Newton BFGS and steepest descent as search direction. A switching criterion is derived based on the First and Second order Kuhn-Tucker condition. The switching criterion can be viewed as a way to change between quasi-Newton and steepest descent step by matching the Kuhn-Tucker condition. This is to ensure that no potential feasible moves away from the current descent step to the other one that reduced the value of the objective function. Numerical results are also presented, which suggest that an improvement has been achieved compared with the BFGS algorithm.

Keywords: Unconstrained optimization, quasi-Newton BFGS, steepest descent, Switching criteria, First and Second order Kuhn-Tucker condition

INTRODUCTION

In this paper, we consider the following unconstrained minimization problem:

$$\min f(x), \text{ where } x\in E^n,\ f\in C^2 \text{ and } f \text{ is strictly convex}. \tag{1.1}$$

In particular, we focus on the class of algorithm known as quasi-Newton or variable-metric methods. These algorithms, which assume the availability of the gradient $g(x)$ for any given x, are based on the recursion

$$x_{k+1} = x_k - \lambda_k H_k g_k . \quad (1.2)$$

In this recursion, λ_k is a positive stepsize parameter selected to satisfy the inexact line search conditions

$$f(x_k + \lambda_k p_k) \leq f(x_k) + \sigma_1 \lambda_k p_k^T \nabla f(x_k) \quad (1.3)$$

$$p_k^T \nabla f(x_k + \lambda_k p_k) \geq \sigma_2 p_k^T \nabla f(x_k) \quad (1.4)$$

where σ_1 and σ_2 are two constants such that $0 < \sigma_1 \leq \sigma_2 < 1$ and $\sigma_1 < 0.5$ and p_k a direction of search, while H_k is an n3 n matrix approximating the inverse Hessian $\left[\nabla^2 f(x)\right]^{-1}$ at the kth. iteration. The objective of these algorithms is to obtain some of the advantages of Newton's method while using only first-order information about the function. Thus, the approximations H_k are inferred from the gradients at previous iterations and updated as new gradients become available. The updating is done such that

$$H_{k+1} \gamma_k = \delta_k \quad (1.5)$$

where

$$\gamma_k = g_{k+1} - g_k \text{ and } \delta_k = x_{k+1} - x_k .$$

This condition, which is often referred to as the quasi-Newton condition, is motivated by the fact that, if the function is quadratic, then

$$\left[\nabla^2 f(x)\right]^{-1} \gamma_k = \delta_k . \quad (1.6)$$

The first algorithm of this type was invented by Davidon (1959). One of the widely accepted formula for the approximation of inverse Hessian is the BFGS formula, briefly discussed in Fletcher (1980).

$$H_{k+1} = H_k + (1 + \gamma_k^T H_k \gamma_k / \delta_k^T \gamma_k)\, \delta_k \delta_k^T / \delta_k^T \gamma_k - (\delta_k \gamma_k^T H + H_k \gamma_k \delta_k^T) / \delta_k^T \gamma_k \quad (1.7)$$

and λ_k is selected such as to minimize $f(x_k - \lambda H_k g_k)$.

However, when one has a poor approximation to an unconstrained minimizer, quick improvement is likely to result from taking a step along the negative gradient (steepest descent). This approach bogs down as the minimum is approached. However, when a good approximate minimizer is available, the quasi-Newton methods often improve it quickly to acceptable accuracy. Considering the above fact, we therefore attempt to improve the quasi-Newton BFGS algorithm by introducing a switching criteria to choose between steepest descent step and Newton step.

PRELIMINARIES AND FORMULATION OF PROBLEM

We first give the following definition :

Definition: A direction p_k is said to be a *descent direction* if there exists a $\lambda_k > 0$ such that

$$f(x_k + \lambda_k p_k) < f(x_k), \qquad \lambda \in (0, \lambda_k). \tag{2.1}$$

A more useful form of the said definition can be written as *Lemma*.

Lemma: A direction p_k is said to be a *descent direction* if there exists a $\lambda_k > 0$ such that

$$g_k^T p_k < 0. \tag{2.2}$$

Proof: We can characterize p_k algebraically by considering the Taylor series expansion of $f(x_k + \lambda_k p_k)$ in terms of $f(x_k)$ and g_k as $\lambda_k \to 0$:

$$f(x_k + \lambda_k p_k) \approx f(x_k p_k) + \lambda_k g_k^T p_k.$$

To satisfy (2.1) with $\lambda_k > 0$ it follows immediately that (2.2) be satisfied.

We can now define a new direction of search d_k, where d_k is the linear combination of the gradient and quasi-Newton directions in the form of

$$d_k = (1-\theta_k)(-H_k g_k) + \theta_k(-g_k), \text{ with } 0 \le \theta_k \le 1. \tag{2.3}$$

The direction d_k equals to the quasi-Newton direction if $\theta_k = 0$ and the steepest descent if $\theta_k = 1$. Furthermore, it is also easy to prove that d_k is a descent direction when H_k is positive-definite. However, when positive-definite BFGS update (1.7) is used, both quasi-Newton and steepest descent directions are descent directions, and one may wonder which direction will yield a better result. We can view the idea of choosing between these two steps as the subproblem of obtaining the optimal θ_k such that

$$\theta_k = \arg\min_{\theta} \; f(x_k + \lambda_k((1-\theta)(-H_k g_k) + \theta(-g_k))),$$

$$\text{subject to} \quad 0 \le \theta_k \le 1 \tag{2.4}$$

where λ_k is the stepsize selected to satisfy certain inexact line search condition. The following assumption is necessary:

Assumption 1. i. θ is independent from λ in every iteration.

ii. We assume that there are no other good approximate minimizers except the Newton and Cauchy point along the line segment from the Newton step to the steepest descent step .

In the following section, we will analyze the necessary and sufficient conditions for constrained minimization subproblem (2.4).

KUHN-TUCKER NECESSARY AND SUFFICIENT CONDITIONS FOR CONSTRAINED MINIMIZATION

Necessary and Sufficient Conditions for a General Constrained Minimization

Consider the general constrained minimization problem :

$$\min \ f(x)$$

$$\text{subject to} \quad e_i(x) = 0 \ , \ i \in I_e \tag{3.1}$$

$$h_i(x) \geq 0 \ , \ i \in I_h$$

where each e_i and h_i is also a real function of x and the set I_e and I_h are composed of the indices of the constraint functions.

A local solution to a constrained optimization problem is found if and only if the variables satisfy the constraints, and no small change to the variables improves the objective function and keeps the constraints satisfied. It follows from this elementary statement that, at the solution to a nonlinear programming problem some fundamental conditions are obtained by the gradient vectors of the objective and constraint functions, including the well-known Kuhn-Tucker conditions. These conditions are stated as theorems below:

Theorem 1: (Kuhn-Tucker Conditions-First-order necessary condition)

If x* is a local solution of the constrained minimization problem, and if the constraint qualification condition is satisfied at x*, then there exist multipliers $\{u_i;\ i \in I_e\}$ and $\{w_i\ ;\ i \in I_h\}$, where each w_i is nonnegative and where w_i is zero if $h_i(x^*)$ is positive, such that the gradient of the objective function at x* has the form

$$\nabla f(x^*) = \sum_{i \in I_e} u_i \nabla e_i(x^*) + \sum_{i \in I_h} w_i \nabla h_i(x^*) \tag{3.2}$$

Theorem 2: (Second-order Necessary Condition)

Let x* be a Kuhn-Tucker point (x* satisfied First-order condition) of the constrained minimization problem, and let the vectors $\{\nabla e_i(x^*)\ ;\ i \in I_e\}$ and $\{\nabla h_i(x^*)\ ;\ i \in I_h\}$ be linearly independent. A necessary condition for x* to be a local solution is that, if s is any vector that satisfied the conditions

$$\begin{aligned} s^T \nabla e_i(x^*) &= 0 \ , \ i \in I_e \\ s^T \nabla h_i(x^*) &= 0 \ , \ i \in I_h \end{aligned} \tag{3.3}$$

then the inequality

$$s^T \nabla^2 L(x^*)\, s \geq 0 \tag{3.4}$$

holds, where

$$L(x) = f(x) - u^T e(x) - w^T h(x) \tag{3.5}$$

is the Lagrangian function with u and w satisfy equation (3.2).

Theorem 3: (Second-order sufficiency Condition)

Let x* be a Kuhn-Tucker point of the constrained minimization problem, and let L(x) be the Lagrangian function (3.5), where u and w are parameters that satisfy expression (3.2). A sufficient condition for x* to be a local solution is that, for every nonzero vector s that satisfies the conditions

$$\begin{aligned} & s^T \nabla f(x^*) \le 0 \\ & s^T \nabla e_i(x^*) = 0 \,,\, i \in I_e \\ & s^T \nabla h_i(x^*) \ge 0 \,,\, i \in I_h \end{aligned} \tag{3.6}$$

the inequality

$$s^T \nabla^2 L(x^*)\, s > 0 \tag{3.7}$$

holds.

The proofs of all three theorems can be found in Powell (1980).

Conditions for Solutions of Subproblem

We can now deviate the conditions for solutions of subproblem (2.4). Let us define a function,

$$\Phi(\theta) = f\left(x_k + \lambda_k\left((1-\theta)(-H_k g_k) + \theta\,(-g_k)\right)\right) \tag{3.8}$$

then , the subproblem (2.4) is equivalent to

$$\theta_k = \arg\min \Phi(\theta)$$

$$\text{subject to} \qquad 0 \le \theta \le 1 \tag{3.9}$$

The linear inequality constraints in subproblem (3.9) are :

$$h_1 = \theta \ge 0 \tag{3.10}$$

and

$$h_2 = 1 - \theta \ge 0. \tag{3.11}$$

The Lagangian function of (3.9):

$$L(\theta) = \Phi(\theta) - w_1\theta - w_2(1-\theta) \tag{3.12}$$

Definition: An inequality constraint is said to be *active* at a point if it is satisfied as an equality there or, in other words, if the point lies on the constraint hypersurface, otherwise the constraint is said to be *inactive* or *passive.*

Inequality constraints (3.10) and (3.11) cannot be active simultaneously. Suppose, at a point θ^* we assume that only inequality constraint h_1 is active. The First-order necessary condition requires that if θ^* is a local solution for (3.9), $w_1 \geq 0$ and $w_2 = 0$ since the inequality constraints h_2 is inactive. Or

$$\frac{d}{d\theta}\left(\Phi\left(\theta^*\right)\right) = w_1 \geq 0 \qquad (3.13)$$

Conversely, if we assume inequality constraint (3.11) is active at θ^* , the First-order necessary conditions give

$$\frac{d}{d\theta}\left(\Phi\left(\theta^*\right)\right) = -w_2 \leq 0 \qquad (3.14)$$

The Second-order conditions state that θ^* is a local solution if $\nabla^2 L(\theta^*)$ is positive definite. However the question remains open if $\nabla^2 L(\theta^*)$ is positive semidefinite. We will give brief discussion why it is sufficient to ignore the positive semidefinite of $\nabla^2 L(\theta^*)$ in the end of this section.

We then only require that

$$\nabla^2 L(\theta^*) = \frac{d^2}{d\theta^2}\left(\Phi\left(\theta^*\right)\right) > 0 \qquad (3.15)$$

If we denote

$$x(\theta) = x^+ + \lambda\ ((1-\theta)(-Hg) + \theta(-g)), \qquad (3.16)$$

in which x^+ is a fixed point along the line on the set of points, $x(\theta)$. By the Chain rule, we have

$$\frac{d}{d\theta} = \sum_i \frac{d}{d\theta} x_i(\theta)\frac{\partial}{\partial x_i} = \lambda(Hg-g)^T\nabla \qquad (3.17)$$

so the *gradient* of $\Phi(\theta)$ $(= f(x(\theta))$ along the line at any point $x(\theta)$ is

$$\frac{d}{d\theta}\Phi(\theta) = \lambda(Hg-g)^T\nabla f = \lambda\nabla f^T(Hg-g) \qquad (3.18)$$

Likewise the *curvature* along the line is

$$\frac{d^2}{d\theta^2}\left(\Phi(\theta)\right) = \lambda^2(Hg-g)^T\nabla^2 f(Hg-g) \qquad (3.19)$$

Then, the Second-order necessary and sufficient conditions require that $\nabla^2 f(x(\theta^*))$ positive semidefinite and positive definite respectively.

Some Argument on Second-order Neccessary Condition

If *theorem 2* is applied to test whether a Kuhn-Tucker point is a local solution of a constrained optimization problem, then we find sometimes that the question is unanswered. We consider this possibility in an important special case, namely when the constraints gradients $\{\nabla e_i(x^*); i\in I_e\}$ and $\{\nabla h_i(x^*); i\in I_h\}$ are linearly independent, and the "strict complementarity" condition is satisfied, which means that, the Lagrangian function (3.5), the multiplier w_i is positive for all i in I_h. The strict complementarity and equation (3.2) imply that a direction s satisfied the conditions (3.3) if and only if it satisfies the conditions (3.6). Therefore, in the special case particularly to our one-dimension problem, the theorem fails to indicate whether a Kuhn-Tucker point is a local solution only if

$$\frac{d^2}{d\theta^2}\left(\Phi(\theta)\right) = \lambda^2 (Hg-g)^T \nabla^2 f(x(\theta^*))(Hg-g) \geq 0 \tag{3.20}$$

holds. At least one of the terms $(Hg-g)^T \nabla^2 f(x(\theta^*))\,(Hg-g)$is zero. This result is an extension of a well-known property of unconstrained minimization calculation; namely that second order conditions for local solutions are inadequate if and only if the second derivative matrix of the objective function is positive semidefinite. A good review for the treatment of this situation is given in Mangasarin (1969).

It is now sufficient to state that any Kuhn-Tucker point θ^* is a local solution if condition (3.15) is satisfied or $\nabla^2 f(x(\theta^*))$ is positive definite, and the positive definiteness of $\nabla^2 f(x(\theta^*))$ is trivial for strictly convex objective function f(x).

We can now summarize our work with *Assumption 1* as follows :

Criteria A. We first consider the Newton point ($\theta_k = 0$),

$$n_{k+1} = x_k - \lambda_k H_k g_k \tag{3.21}$$

If $\nabla f^T(n_{k+1})\,(H_k g_k - g_k) < 0$, the First-order condition is violated. Therefore, with *Assumption 1(ii)*, we reject n_{k+1} and take

$$x_{k+1} = c_{k+1} = x_k - \lambda_k g_k \tag{3.22}$$

Criteria B. We now consider the Cauchy point ($\theta_k = 1$) first,

$$c_{k+1} = x_k - \lambda_k g_k. \tag{3.23}$$

If $\nabla f^T(c_{k+1})(H_k g_k - g_k) > 0$, the First-order condition is violated. Therefore, with *Assumption 1(ii)*, we reject c_{k+1} and take

$$x_{k+1} = n_{k+1} = x_k - \lambda_k H_k g_k. \tag{3.24}$$

ALGORITHMS

Algorithm H1 :

Step 1. Input $H_1 = I$ and x_1, $k := 1$.
If $\|g(x_1)\| \le \varepsilon_1$, stop ; else,

Step2. Compute quasi-Newton direction of search, $p_k = H_k g_k$.

Step 3. Calculate λ_k using inexact line search condition and set Newton point $n_{k+1} = x_k - \lambda_k H_k g_k$.

Step 4. If $(H_k g_k - g_k)^T g(n_{k+1}) \ge 0$, set $x_{k+1} = n_{k+1}$,
update H_{k+1} by (1.6), and go to Step 6,
else, compute steepest descent search direction, $w_k = - g_k$.

Step 5. Calculate λ_k^{new} using inexact line search condition and set Cauchy point, $c_{k+1} = x_k - \lambda_k^{new} g_k$, set $x_{k+1} = c_{k+1}$, and go to Step 6.

Step 6. If $\| g(x_{k+1}) \| \le \varepsilon_1$ or $\| \delta_k \| \le \varepsilon_2$, END,
else, set $k := k+1$, and go to Step 2.

Algorithm H2 :

Step 1. Input $H_1 = I$ and x_1, $k := 1$.
If $\|g(x_1)\| \le \varepsilon_1$, stop ; else,

Step2. Compute $w_k = -g_k$.

Step 3. Calculate λ_k and set $c_{k+1} = x_k - \lambda_k g_k$.

Step 4. If $(H_k g_k - g_k)^T g(c_{k+1}) < 0$, set $x_{k+1} = c_{k+1}$,
update H_{k+1} by (1.6), and go to Step 6,
else, compute $p_k = - H_k g_k$.

Step 5. Calculate λ_k^{new} and set $n_{k+1} = x_k - \lambda_k^{new} H_k g_k$, set $x_{k+1} = n_{k+1}$, and go to Step 6.

Step 6. If $\| g(x_{k+1}) \| \leq \varepsilon_1$ or $\| \delta_k \| \leq \varepsilon_2$, END,
else, set $k := k+1$, and go to Step 2.

RATE OF CONVERGENCE

Under *Preliminaries and formulation of Problem,* we define the descent property

$$g_k^T p_k < 0, \text{ where } p_k \text{ is any descent direction for } \forall k. \tag{5.1}$$

If the search direction p_k is nearly orthogonal to the steepest descent direction $-g_k$, the descent property (5.1) is nearly violated. To exclude this possibility, we assumed that the angle between p_k and $-g_k$ is uniformly bounded away from 90°, that is if

$$\phi_k \leq \pi/2 - \mu \qquad \forall k. \tag{5.2}$$

for some $\mu > 0$, where $\phi_k \in [0, \pi/2]$ is defined by

$$\cos \phi_k = -g_k^T p_k / (\|g_k\|_2 \, \|p_k\|_2) \tag{5.3}$$

We now state a simple global convergence theorem .

Theorem G_1: For a descent method with inexact line search, in which (1.3), (1.4) and (5.2) hold, and if ∇f exists and is uniformly continuous on the level set $\{x: f(x) < f(x_1)\}$, then either $g_k = 0$ for some k, or $f_k(=f(x_k)) \to -\infty$ or $g_k \to 0$.

Proof: Assume that $g_k \neq 0$ for all k (whence $\delta_k \neq 0$) , and that f_k is bounded below; it follows that $f_k - f_{k+1} \to 0$, and hence from (1.2) that $-g_k \delta_k \to 0$. Assume $g_k \to 0$ does not hold. Then $\exists$ an $\varepsilon > 0$ and a subsequence such that $\|g_k\|_2 \geq \varepsilon$ and $\|\delta_k\|_2 \to 0$. Now (5.2) and (5.3) give

$$-g_k \delta_k \geq \sin \mu \, \|g_k\|_2 \, \|\delta_k\|_2 \geq \varepsilon \sin \mu \, \|\delta_k\|_2. \tag{5.4}$$

But a Taylor series gives $f_{k+1} = f_k + g(\xi_k)^T \delta_k$, where ξ_k is on the line segment (x_k, x_{k+1}). By uniform continuity, $g(\xi_k) \to g_k$ as $\delta_k \to 0$, so $g_{k+1}^T \delta_k = g_k \delta_k + o(\|\delta_k\|)$ and $g_{k+1}^T \delta_k / g_k \delta_k \to 1$, which contradicts (1.4). The possibility $f_k \to -\infty$ can usually be eliminated by the nature of $f(x)$.

Assume the sequence x_k (k = 1, 2 ,...) generated by the BFGS algorithm converges to a solution x^* where $\nabla^2 f(x^*)$ is positive definite. If the stepsize $\lambda_k = 1$ is chosen whenever it satisfies the line search condition (1.3) and (1.4), it can be shown that x_k converges to x^* superlinearly, that is

$$\|x_{k+1} - x^*\| / \| x_k - x^*\| \to 0. \tag{5.5}$$

More details can be found in Dennis & Moré (1977).

Now, we can state the convergence theorem for *H1 and H2 Algorithms.*

Theorem G2: Let f satisfies (1.1) on the level set $\{x : f(x) < (x_1)\}$ and let x^* be a locally unique solution to the equation $g(x^*) = 0$. Also assume that $\nabla^2 f$ satisfies a Lipschitz condition at x^*, i.e. $\exists$ a constant κ such that

$$\left\|\nabla^2 f(x) - \nabla^2 f(x^*)\right\| \le \kappa \left\| x - x^* \right\| \tag{5.6}$$

Then, *H1, H2 Algorithms* preserve the local superlinear convergence of the BFGS algorithm and globally convergent when the objective function $f(x)$ is convex, if the inexact line search conditions (1.3) and (1.4) are satisfied.

Proof: Since the direction of search d_k as (2.3) is a descent direction, *theorem G1* guarantees that any descent method (particularly for *H1 and H2 Algorithms* where $\theta = 0$ or 1) converges globally when f is strictly convex. Also when (1.3) and (1.4) are satisfied, for a sequence of $k \to \infty$, $g_k \to 0$, or

$$(H_k g_k - g_k)^T g_{k+1} \to 0 \tag{5.7}$$

or

$$x_{k+1} \to n_{k+1}. \tag{5.8}$$

which preserve the local superlinear convergence of the BFGS algorithm.

NUMERICAL RESULTS

A FORTRAN subroutine was programmed to test the *H1 and H2 Algorithms* presented in the previous section. The following test problems are used.

Problem 1 . Rosenbrook 's function

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

Starting point : i. $(-1.2, 1.0)^T$

ii. $(-12, 10)^T$

iii. $(-120, 100)^T$

Solution : $(1, 1)^T$

Problem 2 . Powell 's function of four variables

$$f(x_1, x_2, x_3, x_4) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$$

Starting point : i. $(3, -1, 0, 1)^T$

ii. $(30, -10, 0, 10)^T$

iii. $(300, -100, 0, 100)^T$

Solution : $(0, 0, 0, 0)^T$

Problem 3 . Wood 's function

$$f(x_1, x_2, x_3, x_4) = 100\,(x_2 - x_1^2)^2 + (1- x_1)^2 + 90\,(x_4 - x_3^2)^2 + (1 - x_3)^2 + 10.1\,[(x_2 - 1)^2 + (x_4 - 1)^2] - 19.8\,(x_2 - 1)\,(x_4 - 1)$$

Starting point : i. $(-3, -1, -3, -1)^T$
ii. $(-30, -10, -30, -10)^T$
iii. $(-150, -50, -150, -50)^T$

Solution : $(1, 1, 1, 1)^T$

Problem 4 . Beale 's function of four variables

$$f(x_1, x_2, x_3, x_4) = \{1.5 - [x_1 (1 - x_1)]\}^2 + \{2.25 - [x_1 (1 - x_2)^2]\}^2 + \{2.65 - [x_1 (1- x_2)^3]\}^2 + \{1.5 - [x_3 (1 - x_4)]\}^2 + \{2.25 - [x_3 (1 - x_4)^2]\}^2 + \{2.65 - [x_3 (1- x_4)^3]\}^2$$

Starting point : i. $(1, 1, 1, 1)^T$
ii. $(5, 5, 5, 5)^T$
iii. $(10, 10, 10, 10)^T$

Solution : $(3, 0.5, 3, 0.5)^T$

The calculation was carried out on an HEWLETT PACKARD *Vectra 486* machine. The convergence criterion is

$$\|g(x_{k+1})\| \le \varepsilon_1 \qquad \text{or} \qquad \|\delta_k\| \le \varepsilon_2. \tag{5.9}$$

For each problem we run for $\varepsilon_1 = 0.0001$ and $\varepsilon_2 = 0.00005$ and the initial matrix H_1 is chosen to be unit matrix *I*. The stepsize $\lambda_k > 0$ satisfying conditions (1.3) and (1.4), with $\sigma_1 \to 0$ and $\sigma_2 \to 1$, or

$$f(x_k + \lambda_k p_k) < f(x_k). \tag{5.10}$$

$$p_k^T \nabla f(x_k + \lambda_k p_k) > p_k^T \nabla f(x_k). \tag{5.11}$$

and, is calculated by cubic approximation with bracketing techniques. More details can be found in Fletcher [3] (1980). All problems are solved, and the numbers of iterations are given in Table 1. Symbol "F" indicates particular algorithm fails. We also solved these problems by the BFGS algorithm, and the numerical results of the BFGS algorithm are also given in Table 1. Table 2 summarizes the results in Table 1.

The numerical results show that the *Algorithm H1* is superior than the original BFGS algorithm on this collection of test problems.

TABLE 1
Comparison of Algorithms H1 and H2 with BFGS

Test Problem	Starting Point	BFGS	H1	H2
1	i.	20	20	F
1	ii.	31	31	28
1	iii.	55	55	56
2	i.	23	21	21
2	ii.	27	25	21
2	iii.	20	20	24
3	i.	32	23	26
3	ii.	34	18	30
3	iii.	39	39	51
4	i.	10	7	10
4	ii.	11	8	10
4	iii.	F	18	20

TABLE 2
Summary of Table 1 - Comparison of H1 and H2 with BFGS

	H1	H2
Superior	7	7
Inferior	0	4
Draw	5	1
Total	12	12

CONCLUSION

In this paper, we have formulated and solved a subproblem to obtain our switching criteria. Incorporating this switching strategy into a quasi-Newton BFGS algorithm, we demonstrated by numerical tests that we can obtain results comparable to the normal quasi-Newton BFGS algorithm. We had chosen the BFGS update which uses formula (1.7) because it is easy to see that the updating *H1* and *H2* algorithms possess the local superlinear and global convergence property of the BFGS algorithm. The numerical results clearly demonstrate that our algorithm is superior than the BFGS algorithm.

REFERENCES

DAVIDON W.C. 1959. Variable metric method for minimization. *Atomic Energy Commision, Research and Development Report No. ANL -5990 (Revised).*

DENNIS J.E. Jr. and J.J. MORÉ. 1977. Quasi-Newton methods, motivation and theory. *SIAM Review* **19(1)**: 46-89.

FLETCHER R. 1980. Practical methods of optimization. In *Unconstrained Optimization.* **1**. New York: John Wiley and Sons.

MANGASARIN O.L. 1969. *Nonlinear Programming.* New York: McGraw-Hill Book Co.

POWELL M. J. D. 1980. Gradient conditions and lagrange multipliers in nonlinear programming. In *Nonlinear Optimization: Theory and Algorithms,* ed. L.C.W. E. Dixon, E. Spedicato and G.P. Szegö, p 202-220. Boston: Birkhäuser.