AMO - Advanced Modeling and Optimization, Volume 11, Number 1, 2009

On Practical Modifications of the Quasi-Newton BFGS Method

M. Al-Baali* and Lucio Grandinetti†

February 23, 2009

Abstract

Recently, several modification techniques have been introduced to the line search BFGS method for unconstrained optimization. These modifications replace the vector of the difference in gradients of the objective function, appearing in the BFGS updating formula, by other modified choices so that certain features are obtained. This paper measures these modifications on the basis of some safeguarded schemes for enforcing the positive definiteness of the Hessian approximations safely in a sense to be defined. Since in the limit the safeguarded conditions are reduced to the Wolfe conditions, the useful theoretical and numerical properties of the BFGS method are maintained. It is shown that some modifications improve the performance of the BFGS method substantially.

Key words. Unconstrained optimization, modified quasi-Newton algorithms, line search framework, Wolfe conditions.

*Department of Mathematics and Statistics, Sultan Qaboos University, Muscat, Oman. (albaali@squ.edu.om)

†Department of Electronics, Informatics and Systems, Calabria University, Rende 87036, Italy. (lugran@unical.it)

AMO - Advanced Modeling and Optimization. ISSN: 1841-4311

1 Introduction

Consider finding a least value of a smooth nonlinear function $f(x) : R^n \to R$ by some modified quasi-Newton algorithms that are defined iteratively in the following way. On every iteration k, an iterate x_k and a positive definite matrix B_k, that approximates the Hessian $G_k = \nabla^2 f(x_k)$, are given and the gradient vector $g_k = \nabla f(x_k)$ and the search direction $s_k = -B_k^{-1} g_k$ are calculated. Then, a steplength α_k is computed such that a new function value f_{k+1}, computed at a new iterate $x_{k+1} = x_k + \alpha_k s_k$, is sufficiently smaller than f_k and usually the curvature condition

$$\delta_k^T \gamma_k > 0, \tag{1.1}$$

where

$$\delta_k = x_{k+1} - x_k, \quad \gamma_k = g_{k+1} - g_k, \tag{1.2}$$

is satisfied. A new Hessian approximation B_{k+1} is obtained by updating B_k in terms of the vectors δ_k and γ_k, which is maintained positive definite if certain quasi-Newton updates, particularly BFGS, are employed and condition (1.1) is satisfied. (For further details, see for example Fletcher, 1987.)

Although the BFGS method is robust, its Hessian approximation has been modified with γ_k replaced by another vector (say, $\widehat{\gamma}_k$) (see for example Yabe, Ogasawara and Yoshino, 2007, and the references therein). Since the proposed γ-modifications seem to approximate $G_{k+1}\delta_k$ 'better' than γ_k in a certain sense, it is expected that the quality of the modified Hessian approximations are improved. Because the positive definiteness property is maintained only if condition (1.1) with γ_k replaced by $\widehat{\gamma}_k$ is satisfied, further *ad hoc* modification is made to ensure that $\widehat{\gamma}_k^T \delta_k > 0$.

This paper, however, shows that substantial improvement can be obtained if γ_k is modified only in certain cases. To distinguish these cases, we will measure the quality of γ_k or its modifications on the basis of the techniques of Powell (1978) and Al-Baali (2003a) for their modified BFGS methods on constrained optimization and nonlinear least squares, respectively. In Section 2, we describe some modified BFGS methods and consider some safeguarded schemes which maintain the useful theoretical and numerical properties that the BFGS method has. Section 3 summarizes some numerical results which we obtained by applying the BFGS algorithm and its modifications to a set of standard test problems. It is shown that the proposed modification techniques improved the performance of the BFGS method substantially. Finally Section 4 concludes the paper. Note that $\|.\|$ will be used to denote the Euclidean norm.

2 Modified BFGS Methods

In the BFGS method, a new Hessian approximation is computed by

$$B_{k+1} = bfgs(B_k, \delta_k, \gamma_k), \tag{2.1}$$

where for any B, δ and γ, the function

$$bfgs(B, \delta, \gamma) = B - \frac{B\delta\delta^T B}{\delta^T B\delta} + \frac{\gamma\gamma^T}{\delta^T\gamma} \tag{2.2}$$

defines the well known BFGS updating formula. Since this formula maintains the positive definiteness of B if the inequality $\delta^T\gamma > 0$ holds, the BFGS Hessian (2.1) is also positive definite provided that the curvature condition (1.1) holds. Indeed this condition is guaranteed by choosing a steplength α_k such that the Wolfe conditions

$$f_k - f_{k+1} \geq \sigma_0 \delta_k^T g_k \tag{2.3}$$

and

$$\delta_k^T \gamma_k \geq -(1 - \sigma_1)\delta_k^T g_k, \tag{2.4}$$

where $\sigma_0 \in (0, 0.5)$ and $\sigma_1 \in (\sigma_0, 1)$, are satisfied. These conditions are used by Powell (1976) to show that the BFGS method converges globally and $q-$superlinearly for convex functions. However the strong Wolfe conditions, defined by (2.3), (2.4) and

$$\delta_k^T \gamma_k \leq -(1 + \sigma_1)\delta_k^T g_k, \tag{2.5}$$

are preferable in practice, because they yield an exact line search with an optimal value of the curvature $\delta_k^T\gamma_k = -\delta_k^T g_k$ if $\sigma_1 = 0$. Since the cost of finding a steplength decreases as σ_1 increases, a sufficiently large value of σ_1 is used (see for example Fletcher, 1987).

For large values of σ_1, say $\sigma_1 \geq 0.9$, an acceptable value of $\delta_k^T\gamma_k$ might be far away from the above optimal value. To rectify this difficulty, Al-Baali (2003b) modifies γ_k to some $\widehat{\gamma}_k$ subject to

$$-(1 - \sigma_1^l)\delta_k^T g_k \leq \widehat{\gamma}_k^T \delta_k \leq -(1 + \sigma_1^u)\delta_k^T g_k, \tag{2.6}$$

where $0 \leq \sigma_1^l < 1$, $\sigma_1^u \geq 0$ (and $\sigma_1^l, \sigma_1^u < \sigma_1$ whenever possible). Note that conditions (2.3) and (2.6) (referred to as Wolfe-like conditions) will be used below and reduced to the strong Wolfe conditions if $\widehat{\gamma}_k = \gamma_k$ and $\sigma_1^l = \sigma_1^u = \sigma_1$. Note that if α_k is chosen only to satisfy the function reduction condition

(2.3), the Wolfe-like conditions can be used to maintain the modified Hesian approximations positive definite.

Since the BFGS method might suffer from large eigenvalues of B_k (see for example Byrd, Liu and Nocedal, 1992, and Powell, 1986) and $\gamma_k = \bar{G}_k \delta_k$, where $\bar{G}_k = \int_0^1 G(x_k + t\delta_k)\, dt$ is the average Hessian matrix along δ_k, γ_k will be modified below if the ratio

$$\rho_k = \frac{\gamma_k^T \delta_k}{\delta_k^T B_k \delta_k} \tag{2.7}$$

is sufficiently far away from one. This ratio seems to measure the quality of γ_k compared to $B_k \delta_k$ reasonably well (see for example Fletcher, 1994, Gill and Leonard, 2003, Al-Baali, Fuduli and Musmanno, 2004, and essentially Powell, 1978).

We now consider some modified techniques for γ_k. Based on the value of the ratio (2.7), Powell (1978) updates γ_k to the hybrid choice

$$\widehat{\gamma}_k = \gamma_k + (1-\varphi)(B_k\delta_k - \gamma_k), \quad \varphi = \frac{0.8}{1-\rho_k}, \tag{2.8}$$

only when $\rho_k < 0.2$ to modify the BFGS update in an SQP method for constrained optimization (further detail can be seen in Fletcher, 1987, and Nocedal and Wright, 1999, for instance).

Al-Baali (2004) extended this modification technique to the limited memory L-BFGS method for unconstrained optimization by enforcing the conditions

$$1 - \sigma_2 \leq \widehat{\rho}_k \leq 1 + \sigma_3, \quad \widehat{\rho}_k = \frac{\widehat{\gamma}_k^T \delta_k}{\delta_k^T B_k \delta_k}, \tag{2.9}$$

where $0 \leq \sigma_2 < 1$ and $\sigma_3 \geq 0$. Note that the choice of steplength $\alpha_k = 1$, which occurs in the limit, reduces (2.9) to the Wolfe-like condition (2.6) which yields that the modified curvature $\widehat{\gamma}_k^T \delta_k$ is sufficiently positive and bounded above. Hence the updated Hessian approximation is computed safely positive definite. Enforcing condition (2.9) with least change in γ_k, the author modifies γ_k to

$$\gamma_k^1 = \begin{cases} \gamma_k + (1-\varphi_k^-)(B_k\delta_k - \gamma_k), & \rho_k < 1 - \sigma_2, \\ \gamma_k + (1-\varphi_k^+)(B_k\delta_k - \gamma_k), & \rho_k > 1 + \sigma_3, \\ \gamma_k, & \text{otherwise}, \end{cases} \tag{2.10}$$

where

$$\varphi_k^- = \frac{\sigma_2}{1-\rho_k}, \quad \varphi_k^+ = \frac{-\sigma_3}{1-\rho_k}. \tag{2.11}$$

Hence $\gamma_k^{1T}\delta_k$ is defined sufficiently close to $\delta_k^T B_k \delta_k$ and the conditioning of the modified-updated matrix B_{k+1} is controlled. We observed that this modification technique improves the performance of the BFGS method substantially (see Section 3, for detail). Therefore we will combine it below with other γ-modifications, described in the following way.

Another technique for modifying γ_k on every iteration is proposed by Zhang, Deng, and Chen (1999) who updated γ_k to

$$\gamma_k^2 = \gamma_k + \frac{t_k}{\|\delta_k\|^2}\delta_k, \tag{2.12}$$

where

$$t_k = 3\Big[2(f_k - f_{k+1}) + (g_{k+1} + g_k)^T\delta_k\Big], \tag{2.13}$$

if $t_k \geq \epsilon_1\|\delta_k\|^2 - \gamma_k^T\delta_k$, for some $\epsilon_1 > 0$. Otherwise, γ_k^2 is given by (2.12) with t_k replaced by the right hand side of the former inequality (which rarely happened for a sufficiently small value of ϵ_1). Hence the inequality $\gamma_k^{2T}\delta_k \geq \epsilon_1\|\delta_k\|^2$ holds on all iterations so that the modified Hessian approximations are maintained positive definite.

The authors showed the useful features of γ_k^2 that it is reduced to γ_k when f is quadratic and for a general sufficiently smooth function f and small $||s_k||$ that

$$(\gamma_k^T\delta_k + t_k) - \delta_k^T G_{k+1}\delta_k = \Big(\gamma_k^T\delta_k - \delta_k^T G_{k+1}\delta_k\Big)O\Big(||\delta_k||\Big). \tag{2.14}$$

This expression yields that $\gamma_k^T\delta_k + t_k (= \gamma_k^{2T}\delta_k)$ approximates $\delta_k^T G_{k+1}\delta_k$ better than $\gamma_k^T\delta_k$. The authors also obtained the global and $q-$superlinear convergence that the BFGS method has for convex functions. Since the corresponding modified method performs worse than the standard BFGS method, other modification techniques have been proposed.

In particular, Zhang and Xu (2001) generalized choice (2.12) to the class of modified vectors

$$\widehat{\gamma}_k = \gamma_k + \frac{t_k}{u^T\delta_k}u, \tag{2.15}$$

where t_k is given by (2.13) and u is any vector such that $u^T\delta_k \neq 0$. Xu and Zhang (2001) extended the above convergence result to class (2.15), provided that $|u^T\delta_k| \geq \epsilon_2\|u\|\|\delta_k\|$ and $\widehat{\gamma}_k^T\delta_k \geq \epsilon_3\gamma_k^T\delta_k$, where $\epsilon_2, \epsilon_3 > 0$ are small numbers. Note that the latter condition will be considered below with $\epsilon_3 = 1 - \sigma_2$.

To maintain the useful invariance linear transformation property of quasi-Newton methods (see for example Fletcher, 1987), the authors considered certain choices for u and recommended $u = \gamma_k$ and the modified vector

$$\tilde{\gamma}_k = \Big(1 + \frac{\tilde{t}_k}{\gamma_k^T \delta_k}\Big)\gamma_k, \tag{2.16}$$

where

$$\tilde{t}_k = \max(t_k, -\sigma_2 \gamma_k^T \delta_k). \tag{2.17}$$

Thus the bound

$$\tilde{\gamma}_k^T \delta_k \geq (1 - \sigma_2)\gamma_k^T \delta_k \tag{2.18}$$

is obtained and, by (2.4), the left side of the Wolfe-like condition (2.6) is satisfied with $\widehat{\gamma}_k$ replaced by $\tilde{\gamma}_k$, $\sigma_1^l = 1 - (1 - \sigma_2)(1 - \sigma_1)$ and $\sigma_1^u = \infty$. Therefore, σ_2 should be chosen so that σ_1^l is sufficiently smaller than one.

For $\sigma_2 = 1 - 10^{-4}$, the authors reported that $\tilde{\gamma}_k$ works better than γ_k, but we observed that this modification worsens the performance of the standard BFGS method in several cases of our experiment. One possible reason for this drawback is that the above choice for σ_2 and the usual choice $\sigma_1 = 0.9$ imply that $\sigma_1^l = 1 - 10^{-5}$ and, by (2.6), that the value of $\tilde{\gamma}_k^T \delta_k$ is not necessarily sufficiently positive. To ensure that $\tilde{\gamma}_k^T \delta_k$ is far away from zero, we tried some values of $\sigma_2 \in (0.01, 0.99)$ and noticed that these choices worsen the performance of (2.16)-(2.17). Therefore, we considered the least value of $\sigma_2 = 0$ which yields $\sigma_1^l = \sigma_1$, but avoids using negative values of t_k. We observed that generally this choice works slightly better than the authors' choice, but improves the performance of the BFGS method substantially when combined with the self-scaling technique in a certain sense (Al-Baali and Khalfan, 2008).

Another possible motivation for the above observations is that when the value of $\tilde{t}_k = -\sigma_2 \gamma_k^T \delta_k$ is used with $\sigma_2 \neq 0$, it follows from (2.16) that $\tilde{\gamma}_k^T \delta_k$ equals neither $\gamma_k^T \delta_k + t_k$ nor $\gamma_k^T \delta_k$. In this case, the useful properties (2.14) of (2.16) with t_k replaced by $\tilde{t}_k$ and that of γ_k are destroyed. To maintain these properties for some nonpositive and bounded values of t_k, we consider the modified vector

$$\gamma_k^3 = \begin{cases} (1 + \dfrac{t_k}{\gamma_k^T \delta_k})\gamma_k, & -\sigma_2 \gamma_k^T \delta_k \leq t_k \leq \sigma_3 \gamma_k^T \delta_k, \\ \gamma_k, & \text{otherwise}, \end{cases} \tag{2.19}$$

where the scalars σ_2 and σ_3 are defined as in (2.9). In practice, we choose values for the latter scalar sufficiently large and for the former one sufficiently far away from both 0 and 1. Thus wide intervals for negative values of t_k and sufficiently positive values of $\gamma_k^{3T} \delta_k$ are obtained. We note that $\gamma_k^{3T} \delta_k \in$

$[1-\sigma_2, 1+\sigma_3](\gamma_k^T\delta_k)$ and (by (2.3)-(2.5)) the Wolfe-like condition (2.6) is satisfied with γ_k^3 replaced by $\widehat{\gamma}_k$ and $\sigma_1^l = 1-(1-\sigma_2)(1-\sigma_1)$ (as for $\tilde{\gamma}_k$) and $\sigma_1^u = (1+\sigma_3)(1+\sigma_1)-1$. For some values of $\sigma_2 \in [0.01, 0.99]$ and of $\sigma_3 \geq 0$, we observed that generally γ_k^3 is preferable to $\tilde{\gamma}_k$.

It is worth noting that the modified BFGS updates, given by (2.1) with γ_k replaced by either γ_k^3 or $\tilde{\gamma}_k$, belong to the Huang (1970) family of updates (see for example Fletcher, 1987) and to the $\gamma-$scaled BFGS update of Biggs (1973) (see for example Zhang and Xu, 2001).

Since the modified choice γ_k^1 retains the invariance property of quasi-Newton methods and works better than the above γ-modifications, it is worth considering class (2.15) with $u = \gamma_k^1$ and maintaining the properties of modification (2.19). Using the interval for t_k as in (2.19), we suggest the modification

$$\gamma_k^4 = \begin{cases} \gamma_k + \dfrac{t_k}{\gamma_k^{1T}\delta_k}\gamma_k^1, & -\sigma_2\gamma_k^T\delta_k \leq t_k \leq \sigma_3\gamma_k^T\delta_k, \\ \gamma_k, & \text{otherwise.} \end{cases} \tag{2.20}$$

This choice (like γ_k^3) improves the quality of γ_k, but worsens that of γ_k^1.

This result can be motivated by the fact that condition (2.9), with $\widehat{\gamma}_k$ replaced by γ_k^m, is guaranteed if $m = 1$. Therefore we consider two further modification techniques to γ_k^m. In one technique, we consider modifying γ_k^m in a manner similar to that of modifying γ_k by formula (2.10). On replacing γ_k by γ_k^m in this formula, it follows that

$$\widehat{\gamma}_k^m = \begin{cases} \gamma_k^m + (1-\widehat{\varphi}_k^-)(B_k\delta_k - \gamma_k^m), & \widehat{\rho}_k^m < 1-\sigma_2, \\ \gamma_k^m + (1-\widehat{\varphi}_k^+)(B_k\delta_k - \gamma_k^m), & \widehat{\rho}_k^m > 1+\sigma_3, \\ \gamma_k^m, & \text{otherwise,} \end{cases} \tag{2.21}$$

where

$$\widehat{\varphi}_k^- = \frac{\sigma_2}{1-\widehat{\rho}_k^m}, \quad \widehat{\varphi}_k^+ = \frac{-\sigma_3}{1-\widehat{\rho}_k^m}, \quad \widehat{\rho}_k^m = \frac{\gamma_k^{mT}\delta_k}{\delta_k^T B_k\delta_k}. \tag{2.22}$$

In practice, this class of modifications improves over γ_k^m for $m = 2, 3, 4$. Note that formula (2.21) reduces to γ_k^1 if $m = 0, 1$, assuming $\gamma_k^0 = \gamma_k$.

Since γ_k might approximate $G_k\delta_k$ or $G_{k+1}\delta_k$ better than $B_k\delta_k$, it is worth maintaining γ_k^m sufficiently close to γ_k. Thus we consider replacing $B_k\delta_k$ by γ_k in class (2.21)-(2.22) so that γ_k^m is updated to

$$\bar{\gamma}_k^m = \begin{cases} \gamma_k^m + (1-\bar{\varphi}_k^-)(\gamma_k - \gamma_k^m), & \bar{\rho}_k^m < 1-\sigma_2, \\ \gamma_k^m + (1-\bar{\varphi}_k^+)(\gamma_k - \gamma_k^m), & \bar{\rho}_k^m > 1+\sigma_3, \\ \gamma_k^m, & \text{otherwise,} \end{cases} \tag{2.23}$$

where

$$\bar{\varphi}_k^- = \frac{\sigma_2}{1 - \bar{\rho}_k^m}, \quad \bar{\varphi}_k^+ = \frac{-\sigma_3}{1 - \bar{\rho}_k^m}, \quad \bar{\rho}_k^m = \frac{\gamma_k^{mT}\delta_k}{\delta_k^T\gamma_k}. \tag{2.24}$$

This class of modifications is proposed by Al-Baali (2003a), in the structured BFGS method for nonlinear least squares, to maintain γ_k^m sufficiently close to γ_k. We note that if $\gamma_k^m = \nu\gamma_k$ (as in (2.16)) and (2.19)), where ν is a scalar, then $\bar{\gamma}_k^m = \hat{\nu}\gamma_k$, for some $\hat{\nu}$, maintains the direction of γ_k. In practice class (2.23) seems to be competitive with class (2.21) for modifying the BFGS method.

We now outline the modified BFGS methods.

Algorithm 3.1: Modified BFGS

0. Given a starting point x_1, a symmetric positive-definite initial Hessian approximation B_1 and values of σ_0 and σ_1. Set $k := 1$.
1. Terminate if a convergence test holds.
2. Compute the search direction $s_k = -B_k^{-1}g_k$.
3. Find a steplength α_k and a new point $x_{k+1} = x_k + \alpha_k s_k$ such that the strong Wolfe conditions

   $$f_{k+1} \leq f_k + \sigma_0\alpha_k g_k^T s_k, \quad |g_{k+1}s_k| \leq -\sigma_1 g_k^T s_k,$$

   (i.e., (2.3)-(2.5)), are satisfied.
4. Compute δ_k and γ_k, defined by (1.2).
5. Choose a modified vector $\widehat{\gamma}_k$.
6. Update B_k by the modified BFGS formula

   $$B_{k+1} = bfgs(B_k, \delta_k, \widehat{\gamma}_k).$$

7. Set $k := k + 1$ and go to Step 1.

Note that, in Step 5, the choice $\widehat{\gamma}_k = \gamma_k$ yields the standard BFGS method (referred to as BFGS_0 in the next section), while a modified choice $\widehat{\gamma}_k$ yields a modified BFGS algorithm.

3 Numerical results

In this section we test the performance of the modified BFGS Algorithm 3.1, using some $\gamma-$modifications considered in this paper. In Step 0, we let the initial Hessian approximation be defined by $B_1 = I$, the unit matrix, and use the values of $\sigma_0 = 10^{-4}$ and $\sigma_1 = 0.9$. The run was stopped in Step 1 when either

$$\|g_k\|^2 \leq \epsilon \max(1, |f_k|),$$

where ϵ is the machine epsilon ($\approx 10^{-16}$), $f_k - f_{k+1} \leq 0$, or the number of iterations reached 10^5. In Step 3, we use Scheme (2.6.4) of Fletcher (1987) for obtaining an acceptable steplength for the strong Wolfe conditions. This scheme is based on some interpolation and firstly tries the initial estimate (2.6.8) of Fletcher, which in the limit becomes 1. In Step 5, when the modified $\widehat{\gamma} = \gamma_k^2$ (given by (2.12)) is used, we let $\epsilon_1 = 10^{-4}$. For the other modifications, we let $\sigma_2 = 0.9$ and $\sigma_3 = 9$.

We study the behaviour of some modified BFGS methods by choosing in Step 5 that $\widehat{\gamma} = \gamma_k^m$, for $m = 0, \ldots, 6$ (the corresponding methods are referred to as BFGS_m), where

- γ_k^0 denotes the standard γ_k.
- γ_k^m, for $m = 1, 3, 4$, are defined by (2.10), (2.19) and (2.20), respectively.
- γ_k^5 denotes $\widehat{\gamma}_k^3$ which modifies γ_k^3 by (2.21)-(2.22) with $m = 3$.
- γ_k^6 denotes $\bar{\gamma}_k^4$ which modifies γ_k^4 by (2.23)-(2.24) with $m = 4$.

We will not report the results for the other γ-modifications considered in this paper, because their performance were worse than that of γ_k^1. Indeed, BFGS_2 performs worse than BFGS_0 (the standard BFGS method).

We implemented the above modified BFGS algorithms in Fortran 77, using Lahey software with double precision arithmetic. We applied these algorithms to a set of 89 standard test problems. The dimensions of 44 tests are small in the range [2,40], while the others are large in the range [100, 400] (for detail, see Table 2 below). One of these tests is proposed by Fletcher and Powell (1963), another can be seen in Grandinetti (1984) and the other tests have been collected and described by Moré, Garbow and Hillstrom (1981) and Conn, Gould and Toint (1988).

To examine the performances of the above modified BFGS algorithms we compared the number of line searches and function and gradient evaluations (referred to as nls, nfe and nge, respectively) required to solve the

Table 1: Modified BFGS Methods

Method	A_l	A_f	A_g
BFGS_1	0.802	0.890	0.805
BFGS_3	0.927	0.928	0.928
BFGS_4	0.937	0.934	0.937
BFGS_5	0.796	0.881	0.802
BFGS_6	0.759	0.813	0.756

test problems with those required by BFGS_0. The numerical results are summarized in Table 1, using the rule of Al-Baali (see for example Al-Baali and Khalfan, 2008). The heading A_l is used to denote the average of certain 89 ratios of nls required to solve the test problems by a method to the corresponding number required by BFGS_0. A value of $A_l < 1$ indicates that the performance of the algorithm compared to that of BFGS_0 improved by $100(1 - A_l)\%$ in terms of nls. Otherwise the algorithm worsens the performance by $100(A_l - 1)\%$. The headings A_f and A_g denote similar ratios with respect to nfe and nge, respectively.

An examination of the results in Table 1 shows that the performance of BFGS_m, for $m = 3, 4$, and for $m = 1, 5, 6$, are respectively a little better than and much better than that of BFGS_0 in terms of nls, nfe and nge.

The improvement of the latter three methods over BFGS_0 is at least 20% in terms of nls and nge and at least 11% in terms of nfe. Although the latter improvement seems small, it is about 40% in terms of the total nfe and nge required to solve all problems and 90% on a few problems. This observation shows a significant improvement of the above three efficient modified BFGS methods over the BFGS_0 method. We also note from Table 1 that BFGS_6 defines the most efficient method with average improvement over BFGS_0 by 24%, 19% and 24% approximately in terms of nls, nfe and nge, respectively.

We now compare BFGS_1 to BFGS_4 which are based on the modified choices γ_k^1 and γ_k^4, respectively. Since the latter modification is also based on the former one and Table 1 shows that BFGS_1 performs much better than BFGS_4, it follows that γ_k^4 worsens γ_k^1. Hence class (2.15) worsens some useful γ-modifications. However, a comparison between BFGS_3 and BFGS_5, which are respectively based on the modified choices γ_k^3 and $\widehat{\gamma}_k^3$, shows that the latter modification improves the former one. Hence our modification technique (2.21) improves some useful choices of class (2.15). Since BFGS_6 depends on the modified choice $\bar{\gamma}_k^4$ which is also based on γ_k^4, the numerical results show that the former modification improves the latter one substantially. Hence the technique (2.23) is also useful in practice.

The above comparisons clearly show that the modified choice γ_k^1 plays an important role for improving the performance of the BFGS method.

4 Conclusion

The numerical results, reported in the previous section, clearly show that our modified classes (2.21) and (2.23) perform well in practice. They are able to improve some modified vectors successfully and yield efficient modified BFGS methods. Although the most efficient BFGS_6 method belongs to class (2.23), we observed that class (2.21) improved several γ-modifications, which are not reported here, better than the former one. Therefore, further numerical experiments are required to choose some typical values for σ_2 and σ_3 to obtain a highly efficient method.

In particular, when the well-known quasi-Newton DFP method was applied to our set of problems, we observed that the modified choice $\widehat{\gamma}_k^3$ works substantially better than $\bar{\gamma}_k^4$, defined by (2.21) and (2.23), for $m = 3$ and $m = 4$, respectively. Indeed the modified choice $\widehat{\gamma}_k^3$ with $\sigma_2 = \sigma_3 = 0.5$ improved the performance of the DFP method significantly with a slight improvement over the standard BFGS method. This support the predictions of Dixon's (1972) result (see for example Fletcher, 1987) that the iterative sequence $\{x_k\}$ of the BFGS and DFP methods is the same in the limit when the strong Wolfe conditions (2.3)-(2.5) are used with sufficiently small values of σ_1.

Acknowledgments

We would like to thank the two anonymous referees for carefully reading a draft of this paper and making a number of valuable comments.

References

[1] Al-Baali, M. 2003a. Quasi-Newton Algorithms for Large-Scale Nonlinear Least-Squares. In High Performance Algorithms and Software for Nonlinear Optimization, Editors G. Di Pillo and A. Murli, Kluwer Academic, pp. 1-21.

[2] Al-Baali, M. 2003b. Quasi-Wolfe Conditions for Newton-Like Methods. *Presented at 18th International Symposium on Mathematical Programming,* Copenhagen, August 18–22.

[3] Al-Baali, M. 2004. Quasi-Wolfe Conditions for Quasi-Newton Methods for Large-Scale Optimization. *Presented at 40th Workshop on Large Scale Nonlinear Optimization,* Erice, Italy, June 22 - July 1.

[4] Al-Baali, M., Fuduli, A. and Musmanno, R. 2004. On the Performance of Switching BFGS/SR1 Algorithms for Unconstrained Optimization. *OMS*, **19**: 153-164.

[5] Al-Baali, M. and Khalfan, H. 2008. A Combined Class of Slef-Scaling and Modified Quasi-Newton Methods, Research Report DOMAS 08/1, Sultan Qaboos University, Oman.

[6] Biggs, M.C. 1973. A Note on Minimization Algorithms which Make Use of Non-Quadratic Properties of the Objective Function. *J. IMA*, **12**: 337–338.

[7] Byrd, R.H., Liu, D.C. and Nocedal, J. 1992. On the Behavior of Broyden's Class of Quasi-Newton Methods. *SIAM J. Optim.*, **2**: 533–557.

[8] Conn, A.R., Gould, N.I.M. and Toint, Ph.L. 1988. Testing a Class of Algorithms for Solving Minimization Problems with Simple Bound on the Variables. *Mathematics of Computation*, **50**: 399–430.

[9] Dixon, L.C.W. 1972. Quasi-Newton algorithms generate identical points, *Math Programming*, **2**: 383–387.

[10] Fletcher, R. 1987. *Practical Methods of Optimization* (2nd edition), Wiley, Chichester, England. (Reprinted in 2000.)

[11] Fletcher, R. 1994. An Overview of Unconstrained Optimization. In Algorithms for Continuous Optimization: The State of the Art, Editor E. Spedicato, Kluwer Academic, pp. 109–143.

[12] Fletcher, R. and Powell, M.J.D. 1963. A Rapidly Convergent Descent Method for Minimization. *Computer J.*, **6:** 163–168.

[13] Gill, P.E., Leonard, M.W. 2003. Limited-Memory Reduced-Hessian Methods for Large-Scale Unconstrained Optimization. *SIAM J. Optim.*, **14:** 380–401.

[14] Grandinetti, L. 1984. Some Investigations in a New Algorithm for Nonlinear Optimization Based on Conic Models of the Objective Function. *JOTA*, **43:** 1–21.

[15] Huang, H.Y. 1970. Unified Approach to Quadratically Convergent Algorithms for Function Minimization. *JOTA*, **5:** 405–423.

[16] Moré J.J., Garbow B.S. and Hillstrom K.E, 1981. Testing Unconstrained Optimization Software. *ACM transactions on mathematical software*, **7**: 17–41.

[17] Nocedal, J. and Wright, S.J. 1999. *Numerical Optimization*, Springer, London.

[18] Powell, M.J.D. 1976. Some Global Convergence Properties of a Variable Metric Algorithm for Minimization without Exact Line Searches. In R.W. Cottle and C.E. Lemke, editors, Nonlinear Programming, SIAM-AMS Proceedings, Vol. IX, SIAM Publications.

[19] Powell, M.J.D. 1978. Algorithms for Nonlinear Constraints that Use Lagrange Functions. *Math. Programming*, **14**: 224–248.

[20] Powell, M.J.D. 1986. How Bad are the BFGS and DFP Methods when the Objective Function is Quadratic? *Math. Programming*, **34**: 34-47.

[21] Xu, C. and Zhang, J. 2001. A Survey of Quasi-Newton Equations and Quasi-Newton Methods for Optimization. *Annals of Operations research*, **103**: 213–234.

[22] Yabe, H., Ogasawara, H. and Yoshino, M. 2007. Local and Superlinear Convergence of Quasi-Newton Methods Based on Modified Secant Conditions. *Journal of Computational and Applied Mathematics*, **205**: 617–632.

[23] Zhang, J.Z., Deng, N.Y. and Chen, L.H. 1999. Quasi-Newton Equation and Related Methods for Unconstrained Optimization. *JOTA*, **102**: 147–167.

[24] Zhang, J.Z. and Xu, C.X. 2001. Properties and Numerical Performance of Quasi-Newton Methods with Modified Quasi-Newton Equations. *J. Comput. Appl. Math.*, **137**: 269–278.

Table 2: Test Functions

Test Code	n	Function's name
MGH3	2	Powell badly scaled
MGH4	2	Brown badly scaled
MGH5	2	Beale
MGH7	3†	Helical valley
MGH9	3	Gaussian
MGH11	3	Gulf research and development
MGH12	3	Box three-dimensional
MGH14	4†	Wood
MGH16	4†	Brown and Dennis
MGH18	6	Biggs Exp 6
MGH20	6,9,12,20	Watson
MGH21	2†,10†,20†, ‡	Extended Rosenbrock
MGH22	4†,12†,20†, ‡	Extended Powell singular
MGH23	10,20, ‡	Penalty I
MGH25	10†,20†, ‡	Variably dimensioned
MGH26	10,20, ‡	Trigonometric of Spedicato
MGH35	8,9,10,20, ‡	Chebyquad
TRIGFP	10,20, ‡	Trigonometric of Fletcher and Powell
CH-ROS	10†,20†, ‡	Chained Rosenbrock
CGT1	8	Generalized Rosenbrock
CGT2	25	Another chained Rosenbrock
CGT4	20	Generalized Powell singular
CGT5	20	Another generalized Powell singular
CGT10	30, ‡	Toint's seven-diagonal generalization of Broyden tridiagonal
CGT11	30, ‡	Generalized Broyden tridiagonal
CGT12	30, ‡	Generalized Broyden banded
CGT13	30, ‡	Another generalized Broyden banded
CGT14	30, ‡	Another Toint's seven-diagonal generalization of Broyden tridiagonal
CGT15	10	Nazareth
CGT16	30, ‡	Trigonometric
CGT17	8, ‡	Generalized Cragg and Levy

†: Two initial points were used; the standard point $\bar{x}$ and $100\bar{x}$.
‡: $n = 40, 100, 200, 400$ were used to define large dimensional tests.
MGH: Tests collected by Moré, Garbow and Hillstrom (1981).
CGT: Tests collected by Conn, Gould and Toint (1988).
TRIGFP: Test given by Fletcher and Powell (1963).
CH-ROS: Test given by Grandinetti (1984).