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)
1 Introduction

Consider finding a least value of a smooth nonlinear function $f(x) : \mathbb{R}^n \to \mathbb{R}$ 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 $\alpha_k$ 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

$$\delta_k^T \gamma_k > 0,$$

where

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

is satisfied. A new Hessian approximation $B_{k+1}$ is obtained by updating $B_k$ in terms of the vectors $\delta_k$ and $\gamma_k$, 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 $\gamma_k$ replaced by another vector (say, $\hat{\gamma}_k$) (see for example Yabe, Ogasawara and Yoshino, 2007, and the references therein). Since the proposed $\gamma$-modifications seem to approximate $G_{k+1}\delta_k$ ‘better’ than $\gamma_k$ 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 $\gamma_k$ replaced by $\hat{\gamma}_k$ is satisfied, further ad hoc modification is made to ensure that $\hat{\gamma}_k^T \delta_k > 0$.

This paper, however, shows that substantial improvement can be obtained if $\gamma_k$ is modified only in certain cases. To distinguish these cases, we will measure the quality of $\gamma_k$ 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 $\|\cdot\|$ will be used to denote the Euclidean norm.
2 Modified BFGS Methods

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

\[ B_{k+1} = bfgs(B_k, \delta_k, \gamma_k), \]  

(2.1)

where for any \( B, \delta \) and \( \gamma \), the function

\[ bfgs(B, \delta, \gamma) = B - \frac{B \delta \delta^T B}{\delta^T B \delta} + \frac{\gamma \gamma^T}{\delta^T \gamma} \]  

(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 \( \alpha_k \) such that the Wolfe conditions

\[ f_k - f_{k+1} \geq \sigma_0 \delta_k^T g_k \]  

(2.3)

and

\[ \delta_k^T \gamma_k \geq -(1 - \sigma_1) \delta_k^T g_k, \]  

(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

\[ \delta_k^T \gamma_k \leq -(1 + \sigma_1) \delta_k^T g_k, \]  

(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 \( \sigma_1 \) increases, a sufficiently large value of \( \sigma_1 \) is used (see for example Fletcher, 1987).

For large values of \( \sigma_1 \), 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 \( \gamma_k \) to some \( \tilde{\gamma}_k \) subject to

\[ -(1 - \sigma_1^l) \delta_k^T g_k \leq \tilde{\gamma}_k \delta_k \leq -(1 + \sigma_1^u) \delta_k^T g_k, \]  

(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 \( \tilde{\gamma}_k = \gamma_k \) and \( \sigma_1^l = \sigma_1^u = \sigma_1 \). Note that if \( \alpha_k \) is chosen only to satisfy the function reduction condition
(2.3), the Wolfe-like conditions can be used to maintain the modified Hessian 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 $\gamma_k = G_k \delta_k$, where $G_k = \int_0^1 G(x_k + t\delta_k) dt$ is the average Hessian matrix along $\delta_k$, $\gamma_k$ will be modified below if the ratio

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

is sufficiently far away from one. This ratio seems to measure the quality of $\gamma_k$ 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 $\gamma_k$. Based on the value of the ratio (2.7), Powell (1978) updates $\gamma_k$ to the hybrid choice

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

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

$$1 - \sigma_2 \leq \tilde{\rho}_k \leq 1 + \sigma_3, \quad \tilde{\rho}_k = \frac{\hat{\gamma}_k^T \delta_k}{\delta_k^T B_k \delta_k}, \quad (2.9)$$

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 $\hat{\gamma}_k^T \delta_k$ is sufficiently positive and bounded above. Hence the updated Hessian approximation is computed safely positive definite. Enforcing condition (2.9) with least change in $\gamma_k$, the author modifies $\gamma_k$ to

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

where

$$\varphi^- = \frac{\sigma_2}{1 - \rho_k}, \quad \varphi^+ = \frac{-\sigma_3}{1 - \rho_k}, \quad (2.11)$$
Hence $\gamma_k^T \delta_k$ 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 $\gamma$-modifications, described in the following way.

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

$$ \gamma_2^k = \gamma_k + \frac{t_k}{\|\delta_k\|^2} \delta_k, \tag{2.12} $$

where

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

if $t_k \geq \epsilon_1 \|\delta_k\|^2 - \gamma_k^T \delta_k$, for some $\epsilon_1 > 0$. Otherwise, $\gamma_2^k$ 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 $\epsilon_1$). Hence the inequality $\gamma_2^T \delta_k \geq \epsilon_1 \|\delta_k\|^2$ holds on all iterations so that the modified Hessian approximations are maintained positive definite.

The authors showed the useful features of $\gamma_2^k$ that it is reduced to $\gamma_k$ when $f$ is quadratic and for a general sufficiently smooth function $f$ and small $\|s_k\|$ that

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

This expression yields that $\gamma_k^T \delta_k + t_k = \gamma_2^T \delta_k$ 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

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

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 $\hat{\gamma}_k^T \delta_k \geq \epsilon_3 \gamma_k^T \delta_k$, 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

\[
\tilde{\gamma}_k = \left( 1 + \frac{\tilde{t}_k}{\gamma_k^T \delta_k} \right) \gamma_k,
\]

where

\[
\tilde{t}_k = \max(t_k, -\sigma_2 \gamma_k^T \delta_k).
\]

Thus the bound

\[
\tilde{\gamma}_k^T \delta_k \geq (1 - \sigma_2) \gamma_k^T \delta_k
\]

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

For \( \sigma_2 = 1 - 10^{-4} \), the authors reported that \( \tilde{\gamma}_k \) works better than \( \gamma_k \), 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 \( \sigma_2 \) and the usual choice \( \sigma_1 = 0.9 \) imply that \( \sigma_1^t = 1 - 10^{-5} \) 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 \( \sigma_1^t = \sigma_1 \), 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 \( \gamma_k \) are destroyed. To maintain these properties for some nonpositive and bounded values of \( t_k \), we consider the modified vector

\[
\gamma_3^T \delta_k = \begin{cases} 
(1 + \frac{t_k}{\gamma_k^T \delta_k}) \gamma_k, & \gamma_k^T \delta_k \leq t_k \leq \sigma_3 \gamma_k^T \delta_k, \\
\gamma_k, & \text{otherwise},
\end{cases}
\]

where the scalars \( \sigma_2 \) and \( \sigma_3 \) 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 \( \gamma_3^T \delta_k \) are obtained. We note that \( \gamma_3^T \delta_k \in \)


\[ 1 - \sigma_2, 1 + \sigma_3\] (\(\gamma_k^T \delta_k\)) and (by (2.3)-(2.5)) the Wolfe-like condition (2.6) is satisfied with \(\gamma_k^3\) replaced by \(\hat{\gamma}_k\) and \(\sigma_1^1 = 1 - (1 - \sigma_2)(1 - \sigma_1)\) (as for \(\hat{\gamma}_k\)) 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 \(\gamma_k^3\) is preferable to \(\hat{\gamma}_k\).

It is worth noting that the modified BFGS updates, given by (2.1) with \(\gamma_k\) replaced by either \(\gamma_k^3\) or \(\hat{\gamma}_k\), 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 \(\gamma_k^4\) retains the invariance property of quasi-Newton methods and works better than the above \(\gamma\)-modifications, it is worth considering class (2.15) with \(u = \gamma_k^4\) and maintaining the properties of modification (2.19). Using the interval for \(t_k\) as in (2.19), we suggest the modification

\[
\gamma_k^4 = \begin{cases} 
\gamma_k + \frac{t_k}{\gamma_k} \gamma_k^1, & -\sigma_2 \gamma_k^T \delta_k \leq t_k \leq \sigma_3 \gamma_k^T \delta_k, \\
\gamma_k, & \text{otherwise.}
\end{cases}
\]

(2.20)

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

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

\[
\gamma_k^m = \begin{cases} 
\gamma_k^m + (1 - \hat{\varphi}_k^-(B_k \delta_k - \gamma_k^m)), & \hat{\rho}_k^m < 1 - \sigma_2, \\
\gamma_k^m + (1 - \hat{\varphi}_k^+(B_k \delta_k - \gamma_k^m)), & \hat{\rho}_k^m > 1 + \sigma_3, \\
\gamma_k^m, & \text{otherwise,}
\end{cases}
\]

(2.21)

where

\[
\hat{\varphi}_k^-= \frac{\sigma_2}{1 - \hat{\rho}_k^m}, \quad \hat{\varphi}_k^+ = \frac{-\sigma_3}{1 - \hat{\rho}_k^m}, \quad \hat{\rho}_k^m = \frac{\gamma_k^m \delta_k}{\delta_k^T B_k \delta_k}.
\]

(2.22)

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

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

\[
\gamma_k^m = \begin{cases} 
\gamma_k^m + (1 - \hat{\varphi}_k^-)(\gamma_k - \gamma_k^m), & \hat{\rho}_k^m < 1 - \sigma_2, \\
\gamma_k^m + (1 - \hat{\varphi}_k^+(\gamma_k - \gamma_k^m)), & \hat{\rho}_k^m > 1 + \sigma_3, \\
\gamma_k^m, & \text{otherwise,}
\end{cases}
\]

(2.23)
where
\[
\bar{\phi}_k^- = \frac{\sigma_2}{1 - \bar{\rho}_k^m}, \quad \bar{\phi}_k^+ = \frac{-\sigma_3}{1 - \bar{\rho}_k^m}, \quad \bar{\rho}_k^m = \frac{\gamma_k^m \delta_k}{\delta_k \gamma_k}.
\] (2.24)

This class of modifications is proposed by Al-Baali (2003a), in the structured BFGS method for nonlinear least squares, to maintain \(\gamma_k^m\) sufficiently close to \(\gamma_k\). We note that if \(\gamma_k^m = \nu \gamma_k\) (as in (2.16)) and (2.19)), where \(\nu\) is a scalar, then \(\bar{\gamma}_k^m = \tilde{\nu} \gamma_k\), for some \(\tilde{\nu}\), maintains the direction of \(\gamma_k\). 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 \(\sigma_0\) and \(\sigma_1\). 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 \(\alpha_k\) 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 \(\delta_k\) and \(\gamma_k\), defined by (1.2).

5. Choose a modified vector \(\tilde{\gamma}_k\).

6. Update \(B_k\) by the modified BFGS formula
\[
B_{k+1} = \text{bfgs}(B_k, \delta_k, \tilde{\gamma}_k).
\]

7. Set \(k := k + 1\) and go to Step 1.

Note that, in Step 5, the choice \(\tilde{\gamma}_k = \gamma_k\) yields the standard BFGS method (referred to as BFGS\(_0\) in the next section), while a modified choice \(\tilde{\gamma}_k\) 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 \( \epsilon \) 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 \( \hat{\gamma} = \gamma_k^2 \) (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 \( \hat{\gamma} = \gamma_k^m \), for \( m = 0, \ldots, 6 \) (the corresponding methods are referred to as BFGS\(_m\)), where

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

We will not report the results for the other \( \gamma \)-modifications considered in this paper, because their performance were worse than that of \( \gamma_k^1 \). 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

<table>
<thead>
<tr>
<th>Method</th>
<th>$A_l$</th>
<th>$A_f$</th>
<th>$A_g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BFGS$_1$</td>
<td>0.802</td>
<td>0.890</td>
<td>0.805</td>
</tr>
<tr>
<td>BFGS$_3$</td>
<td>0.927</td>
<td>0.928</td>
<td>0.928</td>
</tr>
<tr>
<td>BFGS$_4$</td>
<td>0.937</td>
<td>0.934</td>
<td>0.937</td>
</tr>
<tr>
<td>BFGS$_5$</td>
<td>0.796</td>
<td>0.881</td>
<td>0.802</td>
</tr>
<tr>
<td>BFGS$_6$</td>
<td>0.759</td>
<td>0.813</td>
<td>0.756</td>
</tr>
</tbody>
</table>

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 $\gamma_k^1$ and $\gamma_k^4$, 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 $\gamma_k^4$ worsens $\gamma_k^1$. Hence class (2.15) worsens some useful $\gamma$-modifications. However, a comparison between BFGS$_3$ and BFGS$_5$, which are respectively based on the modified choices $\gamma_k^3$ and $\tilde{\gamma}_k^3$, 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 $\tilde{\gamma}_k^4$ which is also based on $\gamma_k^4$,
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 \( \gamma_k^1 \) 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 \( \gamma \)-modifications, which are not reported here, better than the former one. Therefore, further numerical experiments are required to choose some typical values for \( \sigma_2 \) and \( \sigma_3 \) 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 \( \tilde{\gamma}_k^3 \) works substantially better than \( \tilde{\gamma}_k^4 \), defined by (2.21) and (2.23), for \( m = 3 \) and \( m = 4 \), respectively. Indeed the modified choice \( \tilde{\gamma}_k^3 \) 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 \( \sigma_1 \).

### 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


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

†: 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.
TRIGFP: Test given by Fletcher and Powell (1963).