On the Behavior of Damped Quasi-Newton Methods for Unconstrained Optimization

M. Al-Baali1,*, L. Grandinetti2

We consider a family of damped quasi-Newton methods for solving unconstrained optimization problems. This family resembles that of Broyden with line searches, except that the change in gradients is replaced by a certain hybrid vector before updating the current Hessian approximation. This damped technique modifies the Hessian approximations so that they are maintained sufficiently positive definite. Hence, the objective function is reduced sufficiently on each iteration. The recent result that the damped technique maintains the global and superlinear convergence properties of a restricted class of quasi-Newton methods for convex functions is tested on a set of standard unconstrained optimization problems. The behavior of the methods is studied on the basis of the numerical results required to solve these test problems. It is shown that the damped technique improves the performance of quasi-Newton methods substantially in some robust cases (as the BFGS method) and significantly in certain inefficient cases (as the DFP method).

Keywords: Unconstrained optimization, Quasi-Newton methods, Line-search framework.

1. Introduction

We study the behavior of the recent class of damped quasi-Newton methods, proposed by Al-Baali [7] for solving the unconstrained optimization problem

$$\min_{x \in \mathbb{R}} f(x),$$

where $f$ is a nonlinear differentiable function. This damped (D-) class resembles that of Broyden with line searches (see, for example, Dennis and Schnabel [11], Fletcher [12] or Nocedal and Wright [26]) except that the change in gradients $\gamma_k = g_{k+1} - g_k$ is replaced by the hybrid damped-technique

$$\tilde{\gamma}_k = \phi_k \gamma_k + (1-\phi_k) B_k \delta_k, \quad (1)$$

where $\phi_k \in (0,1]$ is a parameter, before updating a Hessian approximation $B_k$. Here, $g_k = \nabla f(x_k)$, $B_k \approx \nabla^2 f(x_k)$, $\delta_k = x_{k+1} - x_k$ and $x_k$ is the current estimate of a solution of the problem.

†Invited paper.
*Corresponding Author.
1 Department of Mathematics and Statistics, Sultan Qaboos University, Muscat, Oman. Email: albaali@squ.edu.om.
2 Department of Electronics, Informatics and Systems, Calabria University, Rende 87036, Italy. Email: lugran@unical.it.
We notice that the value of $\varphi_k = 1$ (or $\tilde{\gamma}_k = \gamma_k$) reduces the damped class of methods to the Broyden family of methods. The latter family contains the standard BFGS and DFP methods, while the former class contains for some $\varphi_k \neq 1$ the corresponding D-BFGS and D-DFP methods, respectively. The D-BFGS method was applied to unconstrained optimization problems for the first time by Al-Baali [4, 5] who extended the D-BFGS method of Powell [21] for constrained optimization in augmented Lagrange and SQP methods (for further detail on the latter case, see for example Fletcher [12] Nocedal and Wright [20] or Gill and Leonard [14].

Although the BFGS method is robust and has several useful numerical and theoretical properties, it suffers from a certain type of ill-conditioned problems. Therefore, several modification techniques have been introduced to the BFGS method see for example Yuan, [25] Zhang et al. [26] Li and Fukushima [16] Xu and Zhang [23] Zhang and Xu [27] Gill and Leonard [14] Al-Baali, [4,5] Wei et al. [22] Yabe et al. [24] Li et al. [17] Al-Baali and Khalfan [8] Al-Baali and Grandinetti [7] and the references therein). Since the latter paper also shows that the above D-technique is preferable to the other modifications for $\gamma_k$ in the BFGS method, here we consider testing not only the D-technique when introduced not only to the inefficient DFP method, but also to other members of the Broyden family of methods. The remainder of our work is organized as follows. In Section 2, we describe the class of damped methods and consider some safeguarded schemes which maintain the useful theoretical and numerical properties of the BFGS method. Section 3 describes some numerical results obtained by applying a selection of methods to a set of standard test problems. It is shown that the proposed damped technique improves the performance of quasi-Newton method substantially in some robust cases (like the BFGS method) and significantly in certain inefficient cases (like the DFP method). Finally, Section 4 gives our concluding remarks. Sections 5 and 6 are appendices.

2. Damped Quasi-Newton Methods

Here we describe the D-Broyden class of quasi-Newton methods. At the beginning of each iteration, a positive definite Hessian approximation $B_k$ is used to define the search direction $s_k$ by solving the system of linear equations $B_k s = -g_k$. Then, a step-length $\alpha_k$ is chosen such that the following Wolfe-Powell conditions hold:

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

and

$$\delta_k^T \gamma_k \geq -(1-\sigma_i) \delta_k^T g_k,$$

where $\sigma_0 \in (0,0.5)$ and $\sigma_i \in (\sigma_0,1)$. Note that the latter inequality ensures that the curvature condition $\delta_k^T \gamma_k > 0$ holds so that the positive definiteness property holds for both the damped and ‘undamped’ Broyden class of methods. For the next iteration, $B_k$ is updated to a new Hessian approximation,

$$B_{k+1} = B_k - \frac{B_k \delta_k \delta_k^T B_k}{\delta_k^T B_k \delta_k} + \frac{\tilde{\gamma}_k \tilde{\gamma}_k^T}{\delta_k^T \tilde{\gamma}_k} + \theta_k \tilde{w}_k \tilde{w}_k^T,$$
\[
\hat{w}_k = \left( \delta_i^T B_k \delta_k \right)^{1/2} \left( \gamma_k \delta_k - \frac{B_k \delta_k}{\delta_i^T B_k \delta_k} \right),
\]
(5)

where \( \gamma_k \) is given by (1), and \( \theta_k \) and \( \varphi_k \) are parameters. This class of damped methods is reduced to the well-known Broyden family of methods, if \( \varphi_k = 1 \) (or \( \gamma_k = \gamma_k \)) for all \( k \), and to the BFGS and DFP methods, if, in addition, \( \theta_k = 0 \) and \( \theta_k = 1 \), respectively. The corresponding damped methods are referred to as D-BFGS and D-DFP, respectively.

\[
\begin{align*}
\sigma_2, & \quad \rho_k < 1 - \sigma_2, \\
\sigma_3, & \quad \rho_k > 1 + \sigma_3, \\
1, & \quad \text{otherwise},
\end{align*}
\]
(6)

where
\[
\rho_k = \frac{\gamma_k^T \delta_k}{\delta_i^T B_k \delta_k},
\]
(7)

\( 0 < \sigma_2 < 1 \), and \( \sigma_3 > 0 \), which is reduced to that of Powell if \( \sigma_2 = 0.8 \) and \( \sigma_3 = \infty \).

For sufficiently small values of \( \sigma_2 \) and \( \sigma_3 \), the choice (6) ensures that \( \gamma_k^T \delta_k \) is sufficiently close to the positive value of \( \delta_i^T B_k \delta_k \) and, hence, the damped formula (4) maintains Hessian approximations sufficiently positive definite. However, the Broyden family satisfies this property only under the restrictions that \( \gamma_k^T \delta_k > 0 \) and \( \theta_k > -1/a_k \), where
\[
a_k = b_k h_k - 1, \quad b_k = \frac{\delta_i^T B_k \delta_k}{\delta_i^T \gamma_k}, \quad h_k = \frac{\gamma_k^T B_k^{-1} \gamma_k}{\delta_i^T \gamma_k},
\]
(8)

Since \( a_k \geq 0 \) (by the Cauchy inequality), the above useful property holds in particular for the nonnegative members \( \theta_k = 0 \) and \( \theta_k = 1 \). Another well-known member of the Broyden family is the symmetric rank 1 (SR1) update, defined by \( \theta_k = 1/(1-b_k) \), which does not belong to the convex class of updates. Since this update does not guarantee the above positive definiteness property and negative values of \( \theta_k \) seem to work well in practice [28], Al-Baali [2] suggested the switching BFGS/SR1 update, given by the non-positive choice
\[
\theta_k = \begin{cases} 
\frac{1}{1-b_k}, & \hat{h}_k < 1, \\
0, & \text{otherwise}.
\end{cases}
\]
(9)
Although the corresponding method of (9) converges globally for convex objective functions, its performance is better than that of the robust BFGS method (see for instance Lukšan and Spedicato [18] and the next section). The latter two choices for \( \theta_k \) also define the damped D-SR1 and D-(BFGS/SR1) updates, respectively. These damped updates satisfy the positive definiteness property for sufficiently small values of \( \phi_k \).

We now outline the damped-Broyden family of quasi-Newton methods.

**Algorithm 2.1. Damped-Broyden Family**

1. **Initialize**
   - Give a starting point \( x_0 \), a symmetric positive-definite initial Hessian approximation \( B_0 \), values of \( \sigma_0 \) and \( \sigma_1 \), and set \( k := 1 \).
   
2. **Terminate**
   - Terminate if a convergence test holds.

3. **Search Direction**
   - Compute the search direction \( s_k = -B_k^{-1}g_k \).

4. **Step-Length**
   - Find a step-length \( \alpha_k \) and a new point \( x_{k+1} = x_k + \alpha_k s_k \) such that the following strong Wolfe-Powell conditions hold:
     \[
     f_{k+1} \leq f_k + \sigma_0 \alpha_k g_k^T s_k, \quad g_k^T s_k \leq -\sigma_1 g_k^T s_k .
     \]

5. **Update**
   - Compute \( \delta_k \), \( \gamma_k \), and \( \rho_k \).

6. **Advanced**
   - Choose values for \( \theta_k \) and \( \varphi_k \) and compute \( \tilde{\gamma}_k \).

7. **Advance**
   - Update \( B_k \) by the D-Broyden formula (4).

8. **Advance**
   - Set \( k := k + 1 \) and go to Step 1.

This algorithm is reduced to the normal Broyden family of methods if the choice \( \varphi_k = 1 \) is used in Step 5 for all iterations (which is also obtained by substituting \( \sigma_2 = 1 \) and \( \sigma_3 = \infty \) into (6)). This choice with, in particular, \( \theta_k = 0 \) yield the standard BFGS method, while \( \varphi_k \neq 1 \), for some \( k \), yields a D-BFGS method. We use in Step 3 the strong Wolfe-Powell conditions, as commonly used in practice, which imply the Wolfe-Powell conditions (2)-(3). Note that Al-Baali [7] also extends the global and superlinear convergence result of Byrd et al. [9], that a restricted Broyden family of methods has for convex functions, to the class of damped methods.

### 3. Numerical Analysis

In this section, we test the performance of Algorithm 2.1 for some values of the updating parameter \( \theta_k \), which was implemented, as in Al-Baali and Grandinetti [7] in Fortran 77, using the Lahey software with double precision arithmetic. In Step 0 of the algorithm, we let the initial Hessian approximation \( B_1 = I \), the identity matrix, and use the values of \( \sigma_0 = 10^{-4} \) and \( \sigma_1 = 0.9 \) in (10). The run was stopped in Step 1 when either

\[
\| g_k \|^2 \leq \epsilon \max \left( 1, |f_k| \right) ,
\]

where \( \epsilon \) is the machine epsilon (\( \approx 10^{-16} \)), \( f_{k+1} \geq f_k \), or the number of iterations reached
In Step 3, we used the scheme \((2.6.4)\) of Fletcher \([12]\) for obtaining an acceptable step-length \(\alpha_k\) for the strong Wolfe-Powell conditions \((10)\). This scheme is based on some function interpolations and firstly tries Fletcher’s initial estimate \((2.6.8)\) for \(\alpha_k\), which is reduced to one in the limit. In Step 5, some values for \(\theta_k\) were considered as below. The default value of \(\phi_k = 1\) was usually used, but for the damped technique is considered we let \(\phi_k\) be defined by formula \((6)\) with several values of \(\sigma_2\) and \(\sigma_3\) chosen on the basis of some results reported in Al-Baali \([7]\). Here, we report the results for the following choices which differ from those considered by Al-Baali and Grandinetti \([7]\) for the D-BFGS method. We let

\[
\sigma_2 = \begin{cases} 
0.5, & \rho_k < 0.5 \text{ and } |\theta_k| \alpha_k > 0.5, \\
\max \left( \min \left( 0.5, \sigma, \nu \right), \nu \right), & \rho_k < 0.5 \text{ and } |\theta_k| \alpha_k > 0.5, \\
1, & \text{otherwise},
\end{cases}
\]

\[(11)\]

where \(\nu = 10^{-7}\) and

\[
\bar{\sigma} = \frac{0.5|1-\rho_k|}{\sqrt{|\theta_k| \alpha_k}}.
\]

\[(12)\]

We also let \(\sigma_3\) be given by \((11)\) and \((12)\) with \(<, 0.5, 1, \text{ and } |\theta_k| \text{ replaced by } >, e, \infty, \text{ and } \max(|\theta_k|, 1), \text{ respectively.}\) The small value of \(\nu = 10^{-7}\) was used to avoid destroying the character of the damped technique. Indeed, this value was never used in our experiments, since we observed that the smallest value for \(\sigma_2\) and \(\sigma_3\) was \(10^{-5}\), which rarely occurred in practice. Thus, we employed the damped technique when the values of the scalars \(|1-\rho_k|\) and \(b_k h_k - 1\) became sufficiently away from zero, because Al-Baali \([7]\) showed that these scalars tend to zero and \(\phi_k \to 1\), when the damped methods converge to the solution superlinearly for convex functions.

To define the parameter \(\theta_k\) in Step 5, we tried several selections for \(\theta_k\). Here, we report the results for the three well known choices of \(\theta_k = 0, \theta_k = 1\) and \((9)\), which maintain the positive definite Hessian approximations. These choices yield the BFGS, DFP and BFGS/SR1 and their corresponding D-BFGS, D-DFP and D-(BFGS/SR1) methods, respectively. We applied these methods (as in Al-Baali and Grandinetti, \([7]\)) to a set of 89 standard test problems, with their names, citations and dimensions (in the range \([2,100]\)) listed in Table 3 in Appendix B.

As expected, the DFP method was inefficient, since it failed to solve about 36% of the test problems and converged very slowly for several other test problems. However, the other methods solved all the test problems successfully.

To examine the behavior of the successful methods, the numerical results are summarized in Tables 1 and 2. Table 1 represents the ratios of the total number of line searches, function evaluations and gradient evaluations required by each method to solve all the test problems in the set to that required by the BFGS method (denoted by \(T_l, T_f\) and \(T_g\),...
respectively). These ratios clearly show that the damped methods are preferable to the
undamped ones. They indicate that the total number of \( l, f \) and \( g \) evaluations required to
solve all the tests in the set by the D-BFGS, D-(BFGS/SR1) and D-DFP methods are at
most 57\%, 62\% and 76\%, respectively, of those required by the BFGS method. Thus, the
damped technique improves the performance of the BFGS method substantially and DFP
method significantly.

Since the ratios in Table 1 do not adequately illustrate the performance of the methods,
we also present Table 2. The column headings \( A_l, A_f \) and \( A_g \) stand for certain ‘average’
ratios related, respectively, to the number of \( l, f \) and \( g \) evaluations required to solve each
test problems by the methods versus those required by the BFGS method, using the fair
rule of Al-Baali (see for example Al-Baali [3] and Appendix B). A value of \( A_l < 1 \)
(similarly for \( A_f \) and \( A_g \)) indicates that the performance of a method compared to that
of BFGS is improved by \( 100 (1 - A_l) \% \) in terms of the number of \( l \).

Although the corresponding ratios for each method in Table 2 are larger than those in
Table 1, these ratios maintain the following observations. The damped technique plays an
important role for improving the performance of robust and inefficient quasi-Newton
methods. We observe that the performance of the D-DFP method is a little better than
the standard BFGS method, the other three methods perform substantially better than
BFGS and D-BFGS is the most efficient method. The latter method performs about 24\%,
17\% and 23\% better than the BFGS method in terms of the number of \( l, f \) and \( g \)
evaluations, respectively. Although BFGS/SR1 performs much better than the BFGS
method, D-(BFGS/SR1) also performs a little better than BFGS/SR1 in terms of \( l \) and
\( g \) and slightly in terms of \( f \). This observation indicates that the damped technique does
not destroy the features of robust methods. We also note that the most efficient D-BFGS
method is slightly better than the D-(BFGS/SR1) method.

<table>
<thead>
<tr>
<th>Table 1. Ratios of total cost as compared to BFGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
</tr>
<tr>
<td>D-BFGS</td>
</tr>
<tr>
<td>D-DFP</td>
</tr>
<tr>
<td>BFGS/SR1</td>
</tr>
<tr>
<td>D-(BFGS/SR1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Average ratios as compared to BFGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
</tr>
<tr>
<td>D-BFGS</td>
</tr>
<tr>
<td>D-DFP</td>
</tr>
<tr>
<td>BFGS/SR1</td>
</tr>
<tr>
<td>D-(BFGS/SR1)</td>
</tr>
</tbody>
</table>
4. Conclusion

We showed that the damped technique works well in practice. The technique improves the performance of inefficient methods significantly and robust methods substantially. The D-BFGS method was recommended, although further experiments are required of finding typical values for $\sigma_2$ and $\sigma_3$ or other useful choices for the damped parameter $\phi$. It is also worth introducing the self-scaling technique to the efficient damped methods in a manner similar to that of Al-Baali and Khalfan [8] who showed that combining the damped and self-scaling techniques yielded a substantial improvement of the BFGS method.

References


Appendix A

We now describe the rule of Al-Baali [1], and also for example [3], for comparing two methods (say M1 and M2) on the basis a set of pair numbers, say \( p_i \) and \( q_i \), for \( i = 1, 2, ..., m \), related to M1 and M2, respectively. In this paper, \( m = 89 \) denotes the number of test problems and both \( p_i \) and \( q_i \) denote, for all \( i \), either the number of line searches, function evaluations or gradient evaluations required to solve a test \( i \) by the M1 and M2 methods, respectively (the latter method is referred to BFGS and the former one to another method under comparison).

Al-Baali modifies the well-known average ratio \( \frac{1}{m} \sum_{i=1}^{m} \frac{p_i}{q_i} \) to the following modified ‘average’ measure,

\[
A_R = \frac{1}{m} \sum_{i=1}^{m} r_i,
\]

where \( r_i \) has the form

\[
r_i = \begin{cases} 
\frac{p_i}{q_i}, & p_i \leq q_i, \\
2 - \frac{p_i}{q_i}, & \text{otherwise.}
\end{cases}
\]

It is assumed that \( r_i = 1 \) if both \( p_i, q_i \to \infty \), which is also used in the following cases. If both M1 and M2 methods either failed or converged to two different solutions, for some test \( i \), then we set \( r_i = 1 \) (i.e., \( p_i = q_i \)). Thus, the \( A_R \) ratio takes all kinds of terminations into account and always belongs to the interval \([0, 2]\). A value of \( A_R \leq 1 \) indicates that the M2 method reduces the cost of (i.e., improves over) M1 by \( 100(1 - A_R)\% \) (or equivalently it is \( 1/A_R \) times better than M1). If \( A_R > 1 \), then M1 is better as in the latter sense but with \( A_R \) replaced by \( 2 - A_R \). Note that if the inequality \( p_i \leq q_i \) holds for all \( i \), then \( A_R \) is reduced to the usual average of the \( m \) ratios \( \left( p_i / q_i \right) \).

Appendix B

Here we present Table 3 consisting of some details on the set of test problems used in this paper. The first column consists of codes and numbers of the tests given in the original sources. One of these tests is proposed by Fletcher and Powell [13] another can be seen in Grandinetti [15] and the other tests have been collected and described by Moré, et al. [19] and Conn, et al. [10]. The second column of the table records the number of variables \( n \) et al. used for each function. We note that the dimensions of 59 test problems range from 2 to 30 and those of the remaining 30 test problems are either 40 or 100. The symbol † indicates that the same test function is used again, but with the initial point multiplied by 100. The third column of the table consists of the function names.
Table 3. The set of test problems

<table>
<thead>
<tr>
<th>Test Code*</th>
<th>Dimension $n$</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†, 40, 100</td>
<td>Extended Rosenbrock</td>
</tr>
<tr>
<td>MGH22</td>
<td>4†, 12†, 20†, 40, 100</td>
<td>Extended Powell singular</td>
</tr>
<tr>
<td>MGH23</td>
<td>10, 20, 40, 100</td>
<td>Penalty I</td>
</tr>
<tr>
<td>MGH25</td>
<td>10†, 20†, 40, 100</td>
<td>Variably dimensioned</td>
</tr>
<tr>
<td>MGH26</td>
<td>10, 20, 40, 100</td>
<td>Trigonometric of Spedicato</td>
</tr>
<tr>
<td>MGH35</td>
<td>8, 9, 10, 20, 40, 100</td>
<td>Chebyquad</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, 40, 100</td>
<td>Toint’s seven-diagonal generalization of Broyden tridiagonal</td>
</tr>
<tr>
<td>CGT11</td>
<td>30, 40, 100</td>
<td>Generalized Broyden tridiagonal</td>
</tr>
<tr>
<td>CGT12</td>
<td>30, 40, 100</td>
<td>Generalized Broyden banded</td>
</tr>
<tr>
<td>CGT13</td>
<td>30, 40, 100</td>
<td>Another generalized Broyden banded</td>
</tr>
<tr>
<td>CGT14</td>
<td>30, 40, 100</td>
<td>Another Toint’s seven-diagonal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of Broyden tridiagonal</td>
</tr>
<tr>
<td>CGT15</td>
<td>10</td>
<td>Nazareth</td>
</tr>
<tr>
<td>CGT16</td>
<td>30, 40, 100</td>
<td>Trigonometric</td>
</tr>
<tr>
<td>CGT17</td>
<td>8, 40, 100</td>
<td>Generalized Cragg and Levy</td>
</tr>
<tr>
<td>CH-ROS</td>
<td>10†, 20†, 40, 100</td>
<td>Chained Rosenbrock</td>
</tr>
<tr>
<td>TRIGFP</td>
<td>10, 20, 40, 100</td>
<td>Trigonometric of Fletcher and Powell</td>
</tr>
</tbody>
</table>

* MGH$m$: Collected by Moré et al. [19] where $m$ denotes the number of the problem test
CGT$m$: Collected by Conn et al. [10] where $m$ denotes the number of the problem test
CH-ROS: Given by Grandinetti [15]
TRIGFP: Given by Fletcher and Powell [13].
†: Two initial points were used; the standard point $\bar{x}$ and 100$\bar{x}$. 