Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 15, No. 1, pp. 45 – 54, 2019

IMPROVED DAMPED QUASI-NEWTON METHODS FOR UNCONSTRAINED OPTIMIZATION

M. Al-Baali and L. Grandinetti

Abstract: Recently, Al-Baali (2014) has extended the damped-technique in the modified BFGS method of Powell (1978) for Lagrange constrained optimization functions to the Broyden family of quasi-Newton methods for unconstrained optimization. Appropriate choices for the damped-parameter, which maintain the global and superlinear convergence property of these methods on convex functions and correct the Hessian approximations successfully, are proposed in this paper.

Key words: *Unconstrained optimization, quasi-Newton methods, damped technique, line search framework*

Mathematics Subject Classification: *90C53, 90C30, 90C46, 65K05*

---

1 Introduction

Consider the recent damped-technique of Al-Baali (2014) - Powell (1978) for improving the behaviour of quasi-Newton algorithms when applied to the unconstrained optimization problem

$$\min_{x \in R^n} f(x). \tag{1.1}$$

It is assumed that $f : R^n \to R$ is a smooth function and its gradient $g(x) = \nabla f(x)$ is computable for all values of x, but its Hessian $G(x) = \nabla^2 f(x)$ may not be available for some x.

Quasi-Newton methods are defined iteratively by $x_{k+1} = x_k - \alpha_k B_k^{-1} g_k$, where α_k is a positive steplength, B_k is a symmetric and positive definite matrix, which approximates the Hessian $G(x_k)$, and $g_k = \nabla f(x_k)$. The Hessian approximation is updated on each iteration to a new B_{k+1} in terms of the difference vectors

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

such that the quasi-Newton condition $B_{k+1} s_k = y_k$ is satisfied. Several formulae for updating B_k have been proposed (see for instance Fletcher, 1987, Dennis and Schnabel, 1996, and Nocedal and Wright, 1999). Here, we consider the one-parameter Broyden family of updates and focus on the well-known BFGS and DFP members which satisfy certain useful properties. In particular, an interval of updates, which contains these members, maintains Hessian approximations positive definite if the new iterate x_{k+1} is chosen such that the curvature condition $s_k^T y_k > 0$ holds. Although the attractive BFGS method has several useful theoretical and numerical properties, it suffers from certain type of ill-conditioned problems (see in particular Powell, 1986). Therefore, several modification techniques have been introduced to the BFGS method to improve its performance (see for example Al-Baali and Grandinetti, 2009, Al-Baali, Spedicato, and Maggioni, 2014, and the references therein).

In this paper we focus on modifying y_k in quasi-Newton updates to the hybrid choice

$$\widehat{y}_k = \varphi_k y_k + (1-\varphi_k) B_k s_k, \tag{1.3}$$

where $\varphi_k \in (0,1]$ is a parameter. This 'damped' parameter is chosen such that the curvature like condition

$$s_k^T \widehat{y}_k > 0 \tag{1.4}$$

holds with a value sufficiently close to $s_k^T B_k s_k$, which is reduced to the curvature condition when $\varphi_k = 1$. A motivation for this modified technique could be stated as follows. Since the curvature condition $s_k^T y_k > 0$ may not hold for the Lagrange constrained optimization function, Powell (1978) suggests the above damped technique for modifying the BFGS update. This technique has been extended by Al-Baali (2014) to all members of the Broyden family of updates for unconstrained optimization.

The resulting two parameters damped (D)-Broyden class of methods and the conditions for obtaining practical global and superlinear convergence result are stated in Section 2. Sections 3 and 4 suggest some modifications to the Powell-AlBaali formula for the damped parameter φ_k, which enforce the convergence property of the D-Broyden class of methods. Section 5 describes some numerical results which shows the usefulness of the damped parameter not only for the Wolfe-Powell , but also for the backtracking line search conditions. Finally, Section 6 concludes the paper.

2 D-Broyden's Class of Methods

Let the Broyden family for updating the current Hessian approximation B_k be given by

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} + \Theta_k w_k w_k^T, \tag{2.1}$$

where Θ_k is a parameter and

$$w_k = (s_k^T B_k s_k)^{1/2} \left(\frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k} \right). \tag{2.2}$$

It is assumed that B_k is symmetric and positive definite and the curvature condition $s_k^T y_k > 0$ holds. This condition is guaranteed by employing the line search framework for computing a new point x_{k+1} such that the Wolfe-Powell conditions

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

and

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

where f_k denotes $f(x_k)$, $\sigma_0 \in (0, 0.5)$ and $\sigma_1 \in (\sigma_0, 1)$, are satisfied. In this case, the Broyden family maintains Hessian approximations positive definite if the updating parameter is chosen such that

$$\Theta_k > \bar{\Theta}_k, \tag{2.5}$$

where

$$\bar{\Theta}_k = \frac{1}{1 - b_k h_k}, \quad b_k = \frac{s_k^T B_k s_k}{s_k^T y_k}, \quad h_k = \frac{y_k^T H_k y_k}{s_k^T y_k} \tag{2.6}$$

and $H_k = B_k^{-1}$. Note that the values of $\Theta_k = 0$ and $\Theta_k = 1$ correspond to the well-known BFGS and DFP updates, respectively. Because $\bar{\Theta}_k < 0$, these values guarantee the positive definiteness property. (For further details see Fletcher, 1987, for instance.)

The D-Broyden class of updates is defined by (2.1) with y_k replaced by $\widehat{y}_k$, given by (1.3). For convenience, this class has been rearranged by Al-Baali (2014) as follows

$$B_{k+1} = B_k + \varphi_k \left(\frac{y_k y_k^T}{s_k^T y_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \phi_k w_k w_k^T \right), \tag{2.7}$$

where

$$\phi_k = \frac{\mu_k}{\varphi_k}(\mu_k \Theta_k + \varphi_k - 1) \tag{2.8}$$

and

$$\mu_k = \frac{\varphi_k}{\varphi_k + (1-\varphi_k) b_k}. \tag{2.9}$$

Thus, in particular, for $\Theta_k = 0$, it follows that $\phi_k < 0$ if $\varphi_k < 1$. Hence, the resulting update (2.7), which is equivalent to the D-BFGS positive definite Hessian approximation

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\widehat{y}_k \widehat{y}_k^T}{s_k^T \widehat{y}_k}, \tag{2.10}$$

has the ability of correcting large eigenvalues of B_k successfully (see for example Al-Baali, 2014, and Byrd, Liu and Nocedal, 1992), unlike the choice of $\varphi_k = 1$ (which corresponds to the usual BFGS update).

In general, we observe that the D-Broyden formula (2.7) maintains the positive definiteness property of Hessian approximations for any choice of Θ_k and sufficiently small values of φ_k, because it yields that $B_{k+1} \to B_k$ as $\varphi_k \to 0$.

Indeed, for well defined values of Θ_k and sufficiently small values of φ_k (or μ_k) which satisfies the inequalities

$$(1-\nu_1)\frac{\bar{\Theta}_k}{\mu_k} \le \mu_k \Theta_k \le 1 - \nu_2, \quad \nu_3 \le \varphi_k \le 1, \tag{2.11}$$

where $\nu_1, \nu_2, \nu_3 > 0$ are preset constants, Al-Baali (2014) extends the global convergence property that the the restricted Broyden family of methods has for convex objective functions to the D-Broyden class of methods. We note that condition (2.11) holds for any well defined choice of Θ_k with sufficiently small values of φ_k, even for $\Theta_k \le \bar{\Theta}_k$ and for $\Theta_k > 1$ which usually yield divergent Broyden methods. This powerful feature of the damped technique has been observed in practice for some choices of Θ_k and φ_k (see Al-Baali, 2014, and Al-Baali and Purnama, 2012).

Al-Baali (2014) also extends the superlinear convergence property that of the Broyden family to one of the D-Broyden class if in addition to condition (2.11) the following condition holds:

$$\sum_{k=1}^{\infty} \ln \left\{ \left(\frac{\varphi_k^2}{\mu_k}\right) \Big[1 + \mu_k^2 \Theta_k (b_k h_k - 1) \Big] \right\} > -\infty. \tag{2.12}$$

The author also shows in the limit that

$$b_k \to 1, \quad b_k h_k \to 1, \quad \varphi_k \to 1. \tag{2.13}$$

Thus when either b_k, $b_k h_k$ and/or their appropriate combinations are sufficiently remote away from one, it might be useful to define $\varphi_k < 1$ which reduces sufficiently the values of the damped scalars $|\widehat{b}_k - 1|$ and $\widehat{b}_k\widehat{h}_k - 1$, where $\widehat{b}_k$ and $\widehat{h}_k$ are equal respectively to b_k and h_k with y_k replaced by $\widehat{y}_k$. We employ this technique in Section 3, using the relations

$$\widehat{b}_k - 1 = \mu_k(b_k - 1), \tag{2.14}$$

$$\widehat{b}_k\widehat{h}_k - 1 = \mu_k^2(b_k h_k - 1) \tag{2.15}$$

which follow by substituting (1.3) after some manipulations (the latter equation is given by Al-Baali, 2014). These relations imply the reductions

$$|\widehat{b}_k - 1| \leq |b_k - 1|, \quad \widehat{b}_k\widehat{h}_k \leq b_k h_k, \tag{2.16}$$

for any μ_k (or φ_k) which belong to the interval $(0, 1]$.

Therefore, for given Θ_k, the damped parameter φ_k should be defined such that condition (2.11) is satisfied, which is possible for an interval of sufficiently small values of φ_k, so that global convergence is obtained. To approach the superlinear convergence, we try to enforce condition (2.12) whenever possible. In the next two sections, we derive some appropriate choices for φ_k and focus on the D-BFGS method which satisfies condition (2.11) for any choice of φ_k and enforces (2.12) if

$$\frac{\varphi_k^2}{\mu_k} \geq 1 \tag{2.17}$$

which holds for sufficiently large values of $\varphi_k < 1$ only if $b_k > 2$ and for $\varphi_k = 1$ without any condition on b_k. The latter values of φ_k should be used near the solution (i.e., by (2.13), when b_k and/or $b_k h_k$ are sufficiently close to one (for further implementation remarks, see Al-Baali, Spedicato, and Maggioni, 2014).

It is worth noting that the above global and superlinear convergence conditions for D-Broyden's class are reduced to those for Broyden's family if $\varphi_k = 1$ is used for all values of k. The analysis for obtaining these conditions is based on that of Byrd, Liu and Nocedal (1992) for Broyden's family with the restricted subclass $\Theta_k \in (\bar{\Theta}_k, 1)$, which extends that of Zhang and Tewarson (1988) for the preconvex subclass $\Theta_k \in (\bar{\Theta}_k, 0)$ with the global convergence property and that of Byrd, Nocedal and Yuan (1987) for the convex subclass $\Theta_k \in [0, 1)$ and Powell (1976) for $\Theta_k = 0$, with the superlinear convergence property, using the result of Dennis and Moré (1974) for the superlinear convergence of quasi-Newton methods.

3 Modifying Powell's Damped Parameter

We now consider finding some choices for the damped parameter φ_k to define the damped vector $\widehat{y}_k$ in (1.3) and hence in the D-Broyden class of updates (2.7). We will focus on the updated choices $\Theta_k = 0$ and $\Theta_k = \frac{1}{1-b_k}$ which correspond to the BFGS and SR1 updates (and their damped updates), respectively, so that the global convergence condition (2.11) is simply satisfied.

Since the scalars b_k and h_k (defined in (2.6)) are undefined if $s_k^T y_k$ is zero or nearly so (which may happen if the second Wolfe-Powell condition (2.4) is not employed), it is preferable to test the well defined reciprocal $\bar{b}_k = 1/b_k$ or $\bar{h}_k = 1/h_k$, where

$$\bar{b}_k = \frac{s_k^T y_k}{s_k^T B_k s_k}, \quad \bar{h}_k = \frac{s_k^T y_k}{y_k^T H_k y_k}. \tag{3.1}$$

Thus, a value of $\bar{b}_k \leq 0$ (or $\bar{h}_k \leq 0$) indicates that y_k should be replaced by $\widehat{y}_k$ with sufficiently small value of φ_k (say, $\varphi_k = 0.9/(1 - \bar{b}_k)$, as in Powell, 1978) so that the curvature like condition (1.4) holds.

To define the first choice of φ_k which maintains the superlinear convergence property, we enforce condition (2.17) which is possible for $\varphi_k \in [\frac{\bar{b}_k}{1-\bar{b}_k}, 1]$ and $\bar{b}_k < 1/2$. In this case, the choice of $\varphi_k = \frac{\sigma_2}{1-\bar{b}_k}$, for $\sigma_2 > 1/2$, can be used. Although condition (2.17) does not hold for $\bar{b}_k > 1/2$, the above replacement of y_k can be used if $\bar{b}_k >> 1$, because it indicates on the basis of the first limit in (2.13) that the iterate is remote away from a solution. In this way, φ_k can be defined as follows

$$\varphi_k^{(1)} = \begin{cases} \dfrac{\sigma_2}{1-\bar{b}_k}, & \bar{b}_k < 1-\sigma_2 \\ \dfrac{\sigma_3}{\bar{b}_k - 1}, & \bar{b}_k > 1+\sigma_3 \\ 1, & \text{otherwise}, \end{cases} \tag{3.2}$$

where $\sigma_2 > 0.5$ and $\sigma_3 \geq e$. This choice with $\sigma_2 = 0.9$ and $\sigma_3 = 9$ (ie, $\varphi_k < 1$ when $\bar{b}_k \notin [0.1, 10]$) is used by Al-Baali and Grandinetti (2009) to define a D-BFGS update, which is reduced to that of Powell (1978) if the latter choice is replaced by $\sigma_3 = \infty$. In the following analyses, it is assumed that $\bar{b}_k > 0$ but otherwise formula (3.2) might be employed.

For an experiment on a simple quadratic function with highly ill-conditioned Hessian, Al-Baali and Purnama (2012) reported that choice (3.2) is not useful enough when $b_k h_k$ is sufficiently close to one. Thus, the authors have added the condition $a_k > \sigma_4$, where

$$a_k = (b_k h_k - 1)\max(|\Theta_k|, 1) \tag{3.3}$$

and $\sigma_4 \geq 0$, to those stated in (3.2). The authors experiment on the quadratic problem shows that the resulting choice with $\Theta_k = 0$ and several values of σ_4 (even for $\sigma_4 = 0$) which define D-BFGS updates work significantly better than both choice (3.2) and the undamped choice $\varphi_k = 1$. However, for general functions and certain values of σ_i, for $i = 0, \dots 4$, which are stated in Section , we observed that the modified damped parameter works a little worse than (3.2). Therefore, we will not consider this modification below, although it improves the performance of the BFGS method substantially.

However, because $a_k > \sigma_4$ is equivalent to both expressions $b_k h_k > 1 + \sigma_4$ and $\bar{b}_k \bar{h}_k < 1 - \sigma_4\, \bar{b}_k \bar{h}_k$, we can eliminate σ_4 and consider the following formula

$$\varphi_k^{(2)} = \begin{cases} \dfrac{\sigma_2}{1-\bar{b}_k}, & \ell_k < 1-\sigma_2 \\ \dfrac{\sigma_3}{\bar{b}_k - 1}, & \ell_k \geq 1-\sigma_2,\ m_k > 1+\sigma_3 \\ 1, & \text{otherwise}, \end{cases} \tag{3.4}$$

where

$$\ell_k = \min(\bar{b}_k, \bar{b}_k \bar{h}_k), \quad m_k = \max(\bar{b}_k, b_k h_k) \tag{3.5}$$

which are smaller and larger than or equal to one, respectively. Note that $\varphi_k^{(2)}$ is reduced to (3.2) if m_k and ℓ_k are replaced by $\bar{b}_k$ in (3.4). It works better than the above damped parameters, although some values of $\varphi_k^{(2)} \notin (0, 1]$ but they are replaced by the undamped choice $\varphi_k^{(2)} = 1$.

Even though, we avoid this case by increasing the size of the interval for the damped parameter as follows

$$\varphi_k^{(3)} = \begin{cases} \dfrac{\sigma_2}{1-\ell_k}, & \ell_k < 1-\sigma_2 \\ \dfrac{\sigma_3}{m_k - 1}, & m_k > 1+\sigma_3 \\ 1, & \text{otherwise} \end{cases} \tag{3.6}$$

which is reduced to (3.2) if m_k and ℓ_k are replaced by $\bar{b}_k$. In general, this choice works well as shown in Section .

$\boxed{4}$ Further Damped Parameters

We now define some choices for the damped parameter φ_k based on the value of $b_k h_k \geq 1$. The first choice has been proposed by Al-Baali and Purnama (2012), that is

$$\varphi_k^{(4)} = \begin{cases} \dfrac{\sigma_4}{\sqrt{a_k}}, & a_k > \sigma_4 \\ 1, & \text{otherwise}, \end{cases} \tag{4.1}$$

where $\sigma_4 > 0$ is a preset constant and a_k is given by (3.3).

This formula is obtained in a manner similar to that used for obtaining (3.2), but on the basis of the second limit in (2.13) and equation (2.15) as follows. If $a_k > \sigma_4$, then we supposed to choose μ_k such that $\widehat{b}_k \widehat{h}_k - 1 = \sigma_4$ which is simply solved, using (2.15), to obtain $\tilde{\mu_k} = \sqrt{\frac{\sigma_4}{a_k}}$. This choice and its corresponding formula of φ_k are considered with other choices by Al-Baali (2014e). However, it is larger or smaller than $\frac{\sigma_4}{\sqrt{a_k}}$ if $\sigma_4 < 1$ or $\sigma_4 > 1$, respectively. Because $\varphi_k \geq \mu_k$ if $\bar{b}_k \leq 1$, we choose $\varphi_k = \frac{\sigma_4}{\sqrt{a_k}}$ if both $\sigma_4 < 1$ and $\bar{b}_k \leq 0.5$ are satisfied so that less changes in y_k is used. However, when $\bar{b}_k > 0.5$ we define $\varphi_k < 1$ only if $\bar{b}_k >> 1$. Therefore, we modify choice (4.1) such that its first case is used when both conditions $a_k > \sigma_4$ and either $\bar{b}_k < 1-\sigma_2$ or $\bar{b}_k > 1+\sigma_3$ are satisfied.

Since the above modified choice works slightly better than (4.1) and similar to that of the BFGS option, we used $\frac{\sigma_4}{\sqrt{a_k}}$ (or replace it by $\sqrt{\frac{\sigma_4}{a_k}}$ to guarantee $\varphi_k \leq 1$) when $1-\sigma_2 \leq \bar{b}_k \leq 1+\sigma_3$ and combined it with choice (3.2) in several ways (see Al-Baali, 2014b). In particular, we let

$$\varphi_k^{(5)} = \begin{cases} \dfrac{\sigma_2}{1-\bar{b}_k}, & \bar{b}_k < 1-\sigma_2 \\ \dfrac{\sigma_3}{\bar{b}_k - 1}, & \bar{b}_k > 1+\sigma_3 \\ \sqrt{\dfrac{\sigma_4}{a_k}}, & 1-\sigma_2 \leq \bar{b}_k \leq 1+\sigma_3, \; a_k > \sigma_4 \\ 1, & \text{otherwise}, \end{cases} \tag{4.2}$$

where $\sigma_4 = \sigma_3$ is used unless otherwise stated. Similarly, combining (4.1) with (3.6), it follows that

$$\varphi_k^{(6)} = \begin{cases} \dfrac{\sigma_2}{1-\ell_k}, & \ell_k < 1-\sigma_2 \\ \dfrac{\sigma_3}{m_k - 1}, & m_k > 1+\sigma_3 \\ \sqrt{\dfrac{\sigma_4}{a_k}}, & \ell_k \geq 1-\sigma_2,\ m_k \leq 1+\sigma_3,\ a_k > \sigma_4 \\ 1, & \text{otherwise,} \end{cases} \tag{4.3}$$

where as above $\sigma_4 = \sigma_3$ is used unless otherwise stated. We observed in practice that both formulae (4.2) and (4.3) work substantially better than choice (4.1) and slightly better than (3.2) and (3.6) (see Section 5 for details).

To involve the value of h_k in computing the damped parameter, we also consider modifying the above choices $\varphi_k^{(2)}, \varphi_k^{(3)}$ and $\varphi_k^{(6)}$ with ℓ_k and m_k replaced by smaller or larger than or equal to values of

$$L_k = \min(\bar{b}_k, \bar{h}_k, \bar{b}_k\bar{h}_k), \quad M_k = \max(\bar{b}_k, \bar{h}_k, b_k h_k), \tag{4.4}$$

respectively. This modification yield a similar performance to the unmodified choices.

$\boxed{5}$ Numerical Experiments

We now test the performance of some members of the D-Broyden class of algorithms which defines the Hessian approximations by (2.7) for $\Theta_k = 0$,

$$\Theta_k = \begin{cases} \dfrac{1}{1-b_k}, & h_k < 0.95 \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

and the choices in the previous sections $\varphi_k = \varphi_k^{(i)}$, for $i = 1, 2, \ldots, 6$, with

$$\sigma_2 = \max(1 - \frac{1}{\alpha_k}, 0.5), \quad \sigma_3 = e, \quad \sigma_4 = 0.95,$$

unless otherwise stated (the latter equation is replaced by $\sigma_4 = \sigma_3$ when $\varphi_k^{(5)}$ and $\varphi_k^{(6)}$ are used). The corresponding classes of D-BFGS and switching D-BFGS/SR1 methods (referred to as $\mathrm{D0}_i$ and $\mathrm{D0S}_i$) are reduced to the attractive undamped BFGS and BFGS/SR1 methods (that $\mathrm{D0}_0$ and $\mathrm{D0S}_0$, respectively) if $\varphi_k = 1$ is used for all values of k. A comparison to the latter two methods is useful, since they work well in practice for the following standard implementation (see for example Al-Baali, 1993, and Lukšan and Spedicato, 2000). For all algorithms, we let the starting Hessian approximation $B_1 = I$, the identity matrix, and compute the steplength α_k such that the strong Wolfe-Powell conditions (2.3), (2.4) and

$$s_k^T y_k \leq -(1+\sigma_1) s_k^T g_k, \tag{5.2}$$

for $\sigma_0 = 10^{-4}$ and $\sigma_1 = 0.9$, are satisfied (based on polynomial interpolations as described for example by Fletcher, 1987, Al-Baali and Fletcher, 1986, and Moré and Thuente, 1994). The iterations were terminated when either $\|g_k\|^2 \leq \epsilon \max(1, |f_k|)$, where ϵ is the machine epsilon ($\approx 10^{-16}$), $f_{k+1} \geq f_k$, or the number of iterations reaches 10^4.

As in Al-Baali (2014), we implemented the above algorithms in Fortran 77, using Lahey software with double precision arithmetic, and applied them to a set of 162 standard test

Table 1: Average ratios of D0i compared to D0$_0$

i	A_l	A_f	A_g
1	0.805	0.856	0.805
2	0.805	0.856	0.805
3	0.803	0.852	0.801
4	1.033	1.048	1.052
5	0.795	0.846	0.796
6	0.803	0.852	0.801

Table 2: Average ratios of D0Si compared to D0$_0$

i	A_l	A_f	A_g
0	0.923	0.942	0.937
1	0.797	0.850	0.795
2	0.797	0.850	0.795
3	0.795	0.850	0.794
4	0.999	1.024	1.026
5	0.786	0.840	0.785
6	0.795	0.850	0.794

problems (most of them belong to CUTEr library and the others are considered by Al-Baali and Grandinetti, 2009, and collected by Andrei, 2008) with n in the range [2,100]. All methods solved the problems successfully.

We compared the number of line searches and function and gradient evaluations (referred to as nls, nfe and nge, respectively, which are required to solve the test problems) to those required by D0$_0$. 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 162 ratios of nls required to solve the test problems by a method to the corresponding number required by the standard BFGS, D0$_0$, method. A value of $A_l < 1$ indicates that the performance of the algorithm compared to that of D0$_0$ 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.

We observe that the performance of the damped D0$_i$ methods, for $i \neq 4$, is substantially better than that of D0$_0$ and D0$_4$ is similar to D0$_0$, in terms of nls, nfe and nge (a similar comparison for D0S$_i$ with D0S$_0$ is also observed from Table 2). Although slight differences among the efficient methods, we observe that D0$_5$ and D0S$_5$ are the winners and the latter one is slightly better than the former one. Even though the tables show that the average improvement of both methods over D0$_0$ are about 20%, 15% and 20% in terms of nls, nfe and nge, respectively, we observed that the reduction of the total of these numbers, which required to solve all problems in the set, is about 40%. Therefore, the damped parameter $\varphi_k^{(5)}$ is recommended in practice.

A comparison of the two tables shows that the performance of the switching D0S$_i$ class of methods is a little better than that of D0$_i$ for each i. Thus the open problem that the former class has the superlinear convergence property that the latter one has for convex functions is illustrated in practice so that it is worth investigating its proof.

Finally it is worth mentioning that the performance of the above efficient damped methods remain better than the standard BFGS method if not only the strong Wolfe-Powell conditions are employed, but also if either the Wolfe-Powell conditions (2.3) and (2.4) are employed or if only the first Wolfe-Powell condition (2.3) is employed. Thus the proposed damped parameters seem appropriate and play an important role for improving the performance of quasi-Newton methods.

6 Conclusion

We have proposed several simple formulae for the damped parameter which maintain the useful theoretical properties of the Broyden class of methods and improve its performance substantially. In particular, they maintain the global and q-superlinear convergence properties, on convex functions, for the standard BFGS and switching BFGS/SR1 methods. The reported numerical results show that the proposed damped parameters are appropriate, since they improve the performance of the standard BFGS method substantially.

Acknowledgements

Presented at the fourth Asian conference on nonlinear analysis and optimization (NAO-Asia), Taipei, Taiwan, August 5 - 9, 2014.

References

[1] M. Al-Baali, Damped techniques for enforcing convergence of quasi-Newton methods, *Optim. Methods Soft.* 29 (2014) 919–936.

[2] M. Al-Baali, New damped quasi-Newton methods for unconstrained optimization, Research Report DOMAS 14/1, Sultan Qaboos University, Oman, 2014a.

[3] M. Al-Baali, Variational quasi-Newton methods for unconstrained optimization, *J. Optim. Theory Appl.* 77 (1993) 127–143.

[4] M. Al-Baali and R. Fletcher, An efficient line search for nonlinear least squares, *J. Optim. Theory Appl.* 48 (1986) 359–378.

[5] M. Al-Baali and L. Grandinetti, On practical modifications of the quasi-Newton BFGS method, *Advanced Modeling Optim.* 11 (2009) 63–76.

[6] M. Al-Baali and A. Purnama, Numerical experience with damped quasi-Newton optimization methods when the objective function is quadratic, Sultan Qaboos University Journal for Science 17 (2012) 1–11.

[7] M. Al-Baali, E. Spedicato and F. Maggioni, Broyden's quasi-Newton methods for nonlinear system of equations and unconstrained optimization: a review and open problems, *Optim. Methods Soft.* 29 (2014) 937–954.

[8] N. Andrei, . An unconstrained optimization test functions collection. *Advanced Modeling Optim.* 10 (2008) 147–161.

[9] R.H. Byrd, D.C. Liu and J. Nocedal, On the behavior of Broyden's class of quasi-Newton methods, *SIAM J. Optim.* 2 (1992) 533–557.

[10] R.H. Byrd, J. Nocedal and Y. Yuan, Global convergence of a class of quasi-Newton methods on convex problems, *SIAM J. Numer. Anal.* 24 (1987) 1171–1190.

[11] J.E. Dennis and J.J. Moré, A characterization of superlinear convergence and its application to quasi-Newton methods, *Math. Comp.* 28 (1974) 549–560.

[12] J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM Publications, 1996.

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

[14] L. Lukšan and E. Spedicato, Variable metric methods for unconstrained optimization and nonlinear least squares, J. Compt. Appl. Math. 124 (2000) 61–95.

[15] J.J. Moré and D.J. Thuente, Line search algorithms with guaranteed sufficient decrease, *ACM Trans. Math. Software* 20 (1994) 286–307.

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

[17] 1M.J.D. Powell, Some global convergence properties of a variable metric algorithm for minimization without exact line searches, in *Nonlinear Programming*, R.W. Cottle and C.E. Lemke (eds.), SIAM-AMS Proceedings, Vol. IX, SIAM Publications, 1976, pp. 53–72.

[18] M.J.D. Powell, Algorithms for nonlinear constraints that use Lagrange functions, *Math. Program.* 14 (1978) 224–248.

[19] M.J.D. Powell, How bad are the BFGS and DFP methods when the objective function is quadratic?, *Math. Program.* 34 (1986) 34–47.

[20] Y. Zhang and R.P Tewarson, . Quasi-Newton algorithms with updates from the preconvex part of Broyden's family, *IMA J. Numer. Anal.* 8 (1988) 487–509.

Manuscript received 17 March 2015
revised 7 September 2015
accepted for publication 16 September 2015

Mehiddin Al-Baali
Department of Mathematics and Statistics
Sultan Qaboos University, Muscat, Oman
E-mail address: albaali@squ.edu.om

Lucio Grandinetti
Department of Electronics, Informatics and Systems
Calabria University, Rende 87036, Italy
E-mail address: lugran@unical.it