

# Improved Damped Quasi-Newton Methods for Unconstrained Optimization\*

Mehiddin Al-Baali<sup>†</sup> and Lucio Grandinetti<sup>‡</sup>

August 2015

## 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

**AMS Subject Classifications.** 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

---

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

<sup>†</sup>Department of Mathematics and Statistics, Sultan Qaboos University, Muscat, Oman. (albaali@squ.edu.om)

<sup>‡</sup>Department of Electronics, Informatics and Systems, Calabria University, Rende 87036, Italy. (lugran@unical.it)

unconstrained optimization problem

$$\min_{x \in R^n} f(x). \quad (1)$$

It is assumed that  $f : R^n \rightarrow R$  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  $\alpha_k$  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

$$s_k = x_{k+1} - x_k, \quad y_k = g_{k+1} - g_k \quad (2)$$

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

$$\hat{y}_k = \varphi_k y_k + (1 - \varphi_k) B_k s_k, \quad (3)$$

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

$$s_k^T \hat{y}_k > 0 \quad (4)$$

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  $\varphi_k$ , which enforce the convergence property of the D-Broyden class of methods. Section 4 describes some numerical results which shows the usefulness of the damped parameter not only for the Wolfe-Powell and backtracking line search conditions. Finally, Section 5 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

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} + \Theta_k w_k w_k^T, \quad (5)$$

where  $\Theta_k$  is a parameter and

$$w_k = (s_k^T B_k s_k)^{1/2} \left( \frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k} \right). \quad (6)$$

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

$$f_k - f_{k+1} \geq -\sigma_0 s_k^T g_k \quad (7)$$

and

$$s_k^T y_k \geq -(1 - \sigma_1) s_k^T g_k, \quad (8)$$

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 maintained Hessian approximations positive definite if the updating parameter is chosen such that

$$\Theta_k > \bar{\Theta}_k, \quad (9)$$

where

$$\bar{\Theta}_k = \frac{1}{1 - b_k h_k}, \quad b_k = \frac{s_k^T B_k s_k}{s_k^T y_k}, \quad h_k = \frac{y_k^T H_k y_k}{s_k^T y_k} \quad (10)$$

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  $\bar{\Theta}_k$ , these values guarantee the positive definiteness property. (For further details see Fletcher, 1987, for instance.)

The D-Broyden class of updates is defined by (5) with  $y_k$  replaced by  $\hat{y}_k$ , given by (3). For convenience, this class has been rearranged by Al-Baali (2014) as follows

$$B_{k+1} = B_k + \varphi_k \left( \frac{y_k y_k^T}{s_k^T y_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \phi_k w_k w_k^T \right), \quad (11)$$

where

$$\phi_k = \frac{\mu_k}{\varphi_k} (\mu_k \Theta_k + \varphi_k - 1) \quad (12)$$

and

$$\mu_k = \frac{\varphi_k}{\varphi_k + (1 - \varphi_k) b_k}. \quad (13)$$

Thus, in particular, for  $\Theta_k = 0$ , it follows that  $\phi_k < 0$  if  $\varphi_k < 1$ . Hence, the resulting update (11), which is equivalent to the D-BFGS positive definite Hessian approximation

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\hat{y}_k \hat{y}_k^T}{s_k^T \hat{y}_k}, \quad (14)$$

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 (11) maintains the positive definiteness property of Hessian approximations for any choice of  $\Theta_k$  and sufficiently small values of  $\varphi_k$ , because it yields that  $B_{k+1} \rightarrow B_k$  as  $\varphi_k \rightarrow 0$ .

Indeed, for well defined values of  $\Theta_k$  and sufficiently small values of  $\varphi_k$  (or  $\mu_k$ ) which satisfies the inequalities

$$(1 - \nu_1) \frac{\bar{\Theta}_k}{\mu_k} \leq \mu_k \Theta_k \leq 1 - \nu_2, \quad \nu_3 \leq \varphi_k \leq 1, \quad (15)$$

where  $\nu_1, \nu_2, \nu_3 > 0$  are preset constants, Al-Baali (2014) extends the global convergence property that the restricted Broyden family of methods has for convex objective functions to the D-Broyden class of methods. We note that condition (15) holds for any well defined choice of  $\Theta_k$  with sufficiently small values of  $\varphi_k$ , even for  $\Theta_k \leq \bar{\Theta}_k$  and  $\Theta_k > 1$  which usually yield divergent Broyden methods. This powerful feature of the damped technique has been observed in practice for some choices of  $\Theta_k$  and  $\varphi_k$  (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 (15) the following condition holds:

$$\sum_{k=1}^{\infty} \ln \left\{ \left( \frac{\varphi_k^2}{\mu_k} \right) [1 + \mu_k^2 \Theta_k (b_k h_k - 1)] \right\} > -\infty. \quad (16)$$

The author also shows in the limit that

$$b_k \rightarrow 1, \quad b_k h_k \rightarrow 1, \quad \varphi_k \rightarrow 1. \quad (17)$$

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  $|\hat{b}_k - 1|$  and  $\hat{b}_k \hat{h}_k - 1$ , where  $\hat{b}_k$  and  $\hat{h}_k$  are equal respectively to  $b_k$  and  $h_k$  with  $y_k$  replaced by  $\hat{y}_k$ . We employ this technique in Section 3, using the relations

$$\hat{b}_k - 1 = \mu_k (b_k - 1), \quad (18)$$

$$\hat{b}_k \hat{h}_k - 1 = \mu_k^2 (b_k h_k - 1) \quad (19)$$

which follow by substituting (3) after some manipulations (the latter equation is given by Al-Baali, 2014). These relations imply the reductions

$$|\hat{b}_k - 1| \leq |b_k - 1|, \quad \hat{b}_k \hat{h}_k \leq b_k h_k, \quad (20)$$

for any  $\mu_k$  (or  $\varphi_k$ ) which belong to the interval  $(0, 1]$ .

Therefore, for given  $\Theta_k$ , the damped parameter  $\varphi_k$  should be defined such that condition (15) is satisfied, which is possible for an interval of sufficiently small values of  $\varphi_k$ , so that global convergence is obtained. To approach the superlinear convergence, we try to enforce condition (16) whenever possible. In the next two sections, we derive some appropriate choices for  $\varphi_k$  and focus

on the D-BFGS method which satisfies condition (15) for any choice of  $\varphi_k$  and enforces (16) if

$$\frac{\varphi_k^2}{\mu_k} \geq 1 \quad (21)$$

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  $\varphi_k$  should be used near the solution (i.e., by (17), 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  $\varphi_k$  to define the damped vector  $\hat{y}_k$  in (3) and hence in the D-Broyden class of updates (11). 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 (15) is simply satisfied.

Since the scalars  $b_k$  and  $h_k$  (defined in (10)) are undefined if  $s_k^T y_k$  is zero or nearly so (which may happen if the second Wolfe-Powell condition (8) 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

$$\bar{b}_k = \frac{s_k^T y_k}{s_k^T B_k s_k}, \quad \bar{h}_k = \frac{s_k^T y_k}{y_k^T H_k y_k}. \quad (22)$$

Thus, a value of  $\bar{b}_k \leq 0$  (or  $\bar{h}_k \leq 0$ ) indicates that  $y_k$  should be replaced by  $\hat{y}_k$  with sufficiently small value of  $\varphi_k$  (say,  $\varphi_k = 0.9/(1 - \bar{b}_k)$ , as in Powell, 1978) so that the curvature like condition (4) holds.

To define the first choice of  $\varphi_k$  which maintains the superlinear convergence property, we enforce condition (21) 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 (21) does not hold for  $\bar{b}_k > 1/2$ , the above replacement of  $y_k$  can be used if  $\bar{b}_k \gg 1$ , because it indicates on the basis of the first limit in (17) that the iterate is remote away from a solution. In this way,  $\varphi_k$  can be defined as follows

$$\varphi_k^{(1)} = \begin{cases} \frac{\sigma_2}{1-\bar{b}_k}, & \bar{b}_k < 1 - \sigma_2 \\ \frac{\sigma_3}{\bar{b}_k - 1}, & \bar{b}_k > 1 + \sigma_3 \\ 1, & \text{otherwise,} \end{cases} \quad (23)$$

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 (23) might be employed.

For an experiment on a simple quadratic function with highly ill-conditioned Hessian, Al-Baali and Purnama (2012) reported that choice (23) 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

$$a_k = (b_k h_k - 1) \max(|\Theta_k|, 1) \quad (24)$$

and  $\sigma_4 \geq 0$ , to those stated in (23). The authors experiment on the quadratic problem shows that the resulting choice with  $\Theta_k = 0$  and several values of  $\sigma_4$  (even for  $\sigma_4 = 0$ ) which define D-BFGS updates work significantly better than both choice (23) and the undamped choice  $\varphi_k = 1$ . However, for general functions and certain values of  $\sigma_i$ , for  $i = 0, \dots, 4$ , which are stated in Section 6, we observed that the modified damped parameter works a little worse than (23). 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  $\sigma_4$  and consider the following formula

$$\varphi_k^{(2)} = \begin{cases} \frac{\sigma_2}{1 - \bar{b}_k}, & \ell_k < 1 - \sigma_2 \\ \frac{\sigma_3}{\bar{b}_k - 1}, & \ell_k \geq 1 - \sigma_2, m_k > 1 + \sigma_3 \\ 1, & \text{otherwise,} \end{cases} \quad (25)$$

where

$$\ell_k = \min(\bar{b}_k, \bar{b}_k \bar{h}_k), \quad m_k = \max(\bar{b}_k, b_k h_k) \quad (26)$$

which are smaller and larger than or equal to one, respectively. Note that  $\varphi_k^{(2)}$  is reduced to (23) if  $m_k$  and  $\ell_k$  are replaced by  $\bar{b}_k$  in (25). 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

$$\varphi_k^{(3)} = \begin{cases} \frac{\sigma_2}{1 - \ell_k}, & \ell_k < 1 - \sigma_2 \\ \frac{\sigma_3}{m_k - 1}, & m_k > 1 + \sigma_3 \\ 1, & \text{otherwise} \end{cases} \quad (27)$$

which is reduced to (23) if  $m_k$  and  $\ell_k$  are replaced by  $\bar{b}_k$ . In general, this choice works well as shown in Section 6.

## 4 Further Damped Parameters

We now define some choices for the damped parameter  $\varphi_k$  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

$$\varphi_k^{(4)} = \begin{cases} \frac{\sigma_4}{\sqrt{a_k}}, & a_k > \sigma_4 \\ 1, & \text{otherwise,} \end{cases} \quad (28)$$

where  $\sigma_4 > 0$  is a preset constant and  $a_k$  is given by (24).

This formula is obtained in a manner similar to that used for obtaining (23), but on the basis of the second limit in (17) and equation (19) as follows. If  $a_k > \sigma_4$ , then we supposed to choose  $\mu_k$  such that  $\hat{b}_k \hat{h}_k - 1 = \sigma_4$

which is simply solved, using (19), to obtain  $\tilde{\mu}_k = \sqrt{\frac{\sigma_4}{a_k}}$ . This choice and its corresponding formula of  $\varphi_k$  are considered with other choices by Al-Baali (2014b). 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  $\bar{b}_k \gg 1$ . Therefore, we modify choice (28) 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 (28) 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 (23) in several ways (see Al-Baali, 2014b). In particular, we let

$$\varphi_k^{(5)} = \begin{cases} \frac{\sigma_2}{1 - \bar{b}_k}, & \bar{b}_k < 1 - \sigma_2 \\ \frac{\sigma_3}{\bar{b}_k - 1}, & \bar{b}_k > 1 + \sigma_3 \\ \sqrt{\frac{\sigma_4}{a_k}}, & 1 - \sigma_2 \leq \bar{b}_k \leq 1 + \sigma_3, a_k > \sigma_4 \\ 1, & \text{otherwise,} \end{cases} \quad (29)$$

where  $\sigma_4 = \sigma_3$  is used unless otherwise stated. Similarly, combining (28) with (27), it follows that

$$\varphi_k^{(6)} = \begin{cases} \frac{\sigma_2}{1 - \ell_k}, & \ell_k < 1 - \sigma_2 \\ \frac{\sigma_3}{m_k - 1}, & m_k > 1 + \sigma_3 \\ \sqrt{\frac{\sigma_4}{a_k}}, & \ell_k \geq 1 - \sigma_2, m_k \leq 1 + \sigma_3, a_k > \sigma_4 \\ 1, & \text{otherwise,} \end{cases} \quad (30)$$

where as above  $\sigma_4 = \sigma_3$  is used unless otherwise stated. We observed in practice that both formulae (29) and (30) work substantially better than choice (28) and slightly better than (23) and (27) (see Section 6 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  $\ell_k$  and  $m_k$  replaced by smaller or larger than or equal to values of

$$L_k = \min(\bar{b}_k, \bar{h}_k, \bar{b}_k \bar{h}_k), \quad M_k = \max(\bar{b}_k, \bar{h}_k, b_k h_k), \quad (31)$$

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

## 5 Numerical Experiments

We now test the performance of some members of the D-Broyden class of algorithms which defines the Hessian approximations by (11) for  $\Theta_k = 0$ ,

$$\Theta_k = \begin{cases} \frac{1}{1 - b_k}, & h_k < 0.95 \\ 0, & \text{otherwise} \end{cases} \quad (32)$$

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

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

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 D0<sub>*i*</sub> and D0S<sub>*i*</sub>) are reduced to the attractive undamped BFGS and BFGS/SR1 methods (that D0<sub>0</sub> and D0S<sub>0</sub>, 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  $\alpha_k$  such that the strong Wolfe-Powell conditions (7), (8) and

$$s_k^T y_k \leq -(1 + \sigma_1) s_k^T g_k, \quad (33)$$

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 Thunete, 1994). The iterations were terminated when either

Table 1: Average ratios of D0*i* compared to D0<sub>0</sub>

<i>i</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 D0S*i* compared to D0<sub>0</sub>

<i>i</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

$\|g_k\|^2 \leq \epsilon \max(1, |f_k|)$ , where  $\epsilon$  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 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<sub>0</sub>. 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<sub>0</sub>, method. A value of  $A_l < 1$  indicates that the performance of the algorithm compared to that of D0<sub>0</sub> 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  $nge$ , we observed that the reduction of the total of these numbers, which require 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 (7) and (8) are employed or if only the first Wolfe-Powell condition (7) 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.

## References

- [1] Al-Baali, M. 2014. Damped Techniques for Enforcing Convergence of Quasi-Newton Methods. *OMS*, **29**: 919–936.
- [2] Al-Baali, M. 2014b. New Damped Quasi-Newton Methods for Unconstrained Optimization. Research Report DOMAS 14/1, Sultan Qaboos University, Oman.
- [3] Al-Baali, M. 1993. Variational Quasi-Newton Methods for Unconstrained Optimization. *JOTA*, **77**: 127–143.
- [4] Al-Baali, M. and Fletcher R. 1986. An Efficient Line Search for Nonlinear Least Squares. *JOTA*, **48**: 359–378.
- [5] Al-Baali, M. and Grandinetti, L. 2009. On Practical Modifications of the Quasi-Newton BFGS Method. *AMO - Advanced Modeling and Optimization*, **11**: 63–76.
- [6] Al-Baali, M. and Purnama, A. 2012. Numerical Experience with Damped Quasi-Newton Optimization Methods When the Objective Function is Quadratic. *Sultan Qaboos University Journal for Science*, **17**: 1–11.
- [7] Al-Baali, M., Spedicato, E., and Maggioni, F. Broyden's Quasi-Newton Methods for Nonlinear System of Equations and Unconstrained Optimization: a Review and Open Problems. *OMS*, **29**: 937-954.
- [8] Andrei, N. 2008. An Unconstrained Optimization Test Functions Collection. *AMO - Advanced Modeling and Optimization*, **10**: 147–161.
- [9] 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.
- [10] Byrd, R.H., Nocedal, J. and Yuan, Y. 1987. Global Convergence of a Class of Quasi-Newton Methods on Convex Problems. *SIAM J. Numer. Anal.*, **24**: 1171–1190.
- [11] Dennis, J.E. and Moré, J.J. 1974. A Characterization of Superlinear Convergence and its Application to Quasi-Newton Methods. *Math. Comp.*, **28**: 549–560.

- [12] Dennis, J.E. and Schnabel, R.B. 1996. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM Publications.
- [13] Fletcher, R. 1987. *Practical Methods of Optimization* (2nd edition), Wiley, Chichester, England. (Reprinted in 2000.)
- [14] Lukšan, L. and Spedicato, E. 2000. Variable Metric Methods for Unconstrained Optimization and Nonlinear Least Squares. *J. Compt. Appl. Math.*, **124**: 61–95.
- [15] Moré, J. J. and Thuente, D. J. 1994. Line Search Algorithms with Guaranteed sufficient decrease. *ACM Trans. Math. Software*, **20**: 286–307.
- [16] Nocedal, J. and Wright, S.J. 1999. *Numerical Optimization*, Springer, London.
- [17] Powell, M.J.D. 1976. Some Global Convergence Properties of a Variable Metric Algorithm for Minimization without Exact Line Searches. In *Nonlinear Programming*, editors R.W. Cottle and C.E. Lemke, SIAM-AMS Proceedings, Vol. IX, SIAM Publications, pp. 53–72.
- [18] Powell, M.J.D. 1978. Algorithms for Nonlinear Constraints that Use Lagrange Functions. *Math. Programming*, **14**: 224–248.
- [19] Powell, M.J.D. 1986. How Bad are the BFGS and DFP Methods when the Objective Function is Quadratic?. *Math. Programming*, **34**: 34–47.
- [20] Zhang, Y. and Tewarson, R.P. 1988. Quasi-Newton Algorithms with Updates from the Preconvex Part of Broyden’s Family. *IMA J. Numer. Anal.*, **8**: 487–509.