Numerical Experience with Damped Quasi-Newton Optimization Methods when the Objective Function is Quadratic

Mehiddin Al-Baali* and Anton Purnama**

Department of Mathematics and Statistics, Sultan Qaboos University, Muscat, Oman, *Email: albaali@squ.edu.om and **Email: antonp@squ.edu.om.

ABSTRACT: A class of damped quasi-Newton methods for nonlinear optimization has recently been proposed by extending the damped-technique of Powell for the BFGS method to the Broyden family of quasi-Newton methods. It has been shown that this damped class has the global and superlinear convergence property that a restricted class of 'undamped' methods has for convex objective functions in unconstrained optimization. To test this result, we applied several members of the Broyden family and their corresponding damped methods to a simple quadratic function and observed several useful features of the damped-technique. These observations and other numerical experiences are described in this paper. The important role of the damped-technique is shown not only for enforcing the above convergence property, but also for improving the performance of efficient, inefficient and divergent undamped methods substantially (significantly in the latter case). Thus, some appropriate ways for employing the damped-technique are suggested.

KEYWORDS: Nonlinear optimization, Quasi-Newton methods, Damped-technique, Line search framework.

خبرة عددية مع تخامد متماثلة نيوتن في الأمثليات عندما تكون دالة الهدف تربيعية

محي الدين البعلي و أنتون برناما

ملخص: تم حديثا اقتراح حزمة طرق متخامدة لمتماثلة نيوتن في الأمثليات غير الخطية بتطوير تقنية باول إلى حزمة برويدن لطرق متماثلة نيوتن. وقد تم إثبات أن هذه الحزمة المتخامدة تمتلك خاصية التقارب الخطي الفائق التي تمتلكها الحزمة المحددة للطرق غير المتخامدة في الأمثليات غير المقيدة عندما تكون دوال الهدف محدبة. لاختبار هذه النتيجة، طبقنا عدداً من عناصر حزمة برويدن وما يقابلها من الطرق المتخامدة على دالة تربيعية بسيطة ولاحظنا عدداً من الصفات الجيدة لتقنية التخامد. وقد تم شرح هذه الملاحظات وغيرها من التجارب العددية الأخرى في هذه الورقة. حيث تم توضيح الدور الرئيسي لتقنية التخامد ليس من أجل الحصول على التقارب فحسب، بل لتحسين سير الطرق السريعة والبطيئة وغير المتقاربة بشكل كبير (ومتميز في الحالة الأخيرة). لذا تم اقتراح طرق مناسبة لاستخدام تقنية التخامد.

MEHIDDIN AL-BAALI and ANTON PURNAMA

1. Introduction

Consider iterative quasi-Newton methods with line search for solving the unconstrained optimization problem

$$\min f(x),$$

where $f : R^n \to R$ is a smooth function.

On each iteration k of these methods, an estimate of a solution x_k and a positive definite Hessian approximation B_k are used to obtain a new estimate x_{k+1}. Then, B_k is updated to a new Hessian approximation B_{k+1} by using a member of the Broyden family in terms of the differences in points $s_k = x_{k+1} - x_k$ and gradients $y_k = g_{k+1} - g_k$, where g_k denotes the gradient $\nabla f(x_k)$. This family consists of efficient, inefficient and divergent methods. The former two types of methods are usually defined sufficiently closely to the well known BFGS and DFP methods, respectively, while the latter type is defined sufficiently remotely away from these methods. For further details see for instance (Fletcher, 1987; Dennis and Schnabel, 1996; Nocedal and Wright, 1999).

Al-Baali and Grandinetti (2009) show that the performance of the BFGS method can be improved if y_k is modified before updating to the damped-technique

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

where φ_k is a parameter chosen appropriately and sufficiently large in the interval $(0,1]$. The resulting damped (D)-BFGS method is proposed by Powell (1978) for the Lagrangian function in constrained optimization and used many times with only values of $\varphi_k \geq 0.8$, see for example (Fletcher, 1987; Nocedal and Wright, 1999). The aim of this paper is to show that small values of φ_k can be used to improve the behaviour of not only the BFGS method, but also all members of the Broyden family of methods for unconstrained optimization. To illustrate this possibility, we applied several members of this family and their corresponding damped methods to a simple quadratic function of two variables, using certain initial point x_1 and Hessian approximation B_1. For particular choices of φ_k, it is shown that many damped methods work substantially better than the BFGS method. The analysis is organized in the following way. Section 2 describes the Broyden and D-Broyden classes of methods, Section 3 provides some numerical results and Section 4 concludes the paper.

2. Damped quasi-Newton methods

We now briefly describe the Broyden and D-Broyden classes of methods. On each iteration of these methods, the current point x_k and a symmetric and positive-definite Hessian approximation B_k are available, given initially x_1 and B_1. Using these available data, a new point $x_{k+1} = x_k - \alpha_k B_k^{-1} g_k$ is calculated. Here, α_k denotes a steplength which is usually chosen such that the Wolfe conditions

$$f_k - f_{k+1} \geq -\sigma_0 s_k^T g_k \tag{2}$$

and

$$s_k^T y_k \geq -(1 - \sigma_1) s_k^T g_k, \tag{3}$$

where f_k denotes $f(x_k)$, $\sigma_0 \in (0, 0.5)$ and $\sigma_1 \in (\sigma_0, 1)$, are satisfied. Then B_k is updated to a new Hessian approximation

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\hat{y}_k \hat{y}_k^T}{s_k^T \hat{y}_k} + \Theta_k (s_k^T B_k s_k) \hat{v}_k \hat{v}_k^T, \tag{4}$$

2

where Θ_k is a parameter,

$$\hat{v}_k = \frac{\hat{y}_k}{s_k^T \hat{y}_k} - \frac{B_k s_k}{s_k^T B_k s_k} \tag{5}$$

and $\hat{y}_k$ is defined by (1) for a suitable value of φ_k. This class of damped updates is reduced to the Broyden family if $\hat{y}_k = y_k$ (which corresponds to $\varphi_k = 1$). Thus, if this equality holds for all iterations we obtain the Broyden family of methods. Otherwise, we obtain the D-Broyden class of methods. In particular, the choices $\Theta_k = 0$ and $\Theta_k = 1$ yield the D-BFGS and D-DFP methods which correspond (for these choices and $\varphi_k = 1$) to the well known BFGS and DFP methods, respectively.

Although the restricted class of quasi-Newton methods (defined for $\varphi_k = 1$ and $\xi_k < \Theta_k < 1$, where ξ_k is a certain negative value sufficiently close to zero) converges globally and q-superlinearly for convex functions, the performance of the class becomes worse as Θ_k increases. In general, the BFGS method is robust and the DFP method is inefficient, though the former method usually suffers from large eigenvalues of B_k, see in particular (Powell, 1986; Byrd *et al.*, 1992). Therefore several modification techniques have been introduced to the BFGS method, see for example (Yabe *et al.*, 2007; Gratton and Toint, 2010). Here, we test the D-Broyden class of methods which is an extension of the D-BFGS update of Powell (1978) in the augmented Lagrangian and SQP methods for constrained optimization. For details on the latter case, see for instance (Fletcher, 1987; Nocedal and Wright, 1999). This damped method was applied to unconstrained optimization problems for the first time by Al-Baali (2003) who considered

$$\varphi_k = \begin{cases} \dfrac{\sigma_2}{1-\rho_k}, & \rho_k < 1-\sigma_2 \\[2mm] \dfrac{\sigma_3}{\rho_k - 1}, & \rho_k > 1+\sigma_3 \\[2mm] 1, & \text{otherwise}, \end{cases} \tag{6}$$

where

$$\rho_k = \frac{s_k^T y_k}{s_k^T B_k s_k} \tag{7}$$

and $\sigma_2, \sigma_3 \geq 0$. If $\sigma_3 \geq (\rho_k - 1)c$, where $c > 0$, then the bound $\varphi_k \geq c$ is guaranteed. This formula with damped-technique (1) ensures that $s_k^T \hat{y}_k$ is sufficiently close to either the curvature $s_k^T y_k$ or approximate curvature $s_k^T B_k s_k$ if φ_k is sufficiently large or small, respectively. Another useful feature of formula (6) is that $s_k^T \hat{y}_k$ remains sufficiently positive even when $s_k^T y_k \leq 0$ (which may occur if the Wolfe condition (3) is not enforced on s_k).

We also observe that formula (6) is reduced to the choice of Powell (1978) if $\sigma_2 = 0.8$ and $\sigma_3 = \infty$. For certain choices of these parameters, Al-Baali (2004) introduced formula (6) to the limited memory L-BFGS method of Nocedal (1980) and reported encouraging numerical results in certain cases. Other encouraging results for the D-BFGS method with formula (6) on unconstrained optimization have been reported by Al-Baali and Grandinetti (2009). The authors also show that this formula works better than other modified BFGS formulae which modify y_k to the form

$$\hat{y}_k = y_k + [6(f_k - f_{k+1}) + 3(g_{k+1} + g_k)^T s_k] u_k, \tag{8}$$

for certain vectors , see for example (Yabe *et al.*, 2007) and the references therein. Because this form is reduced to y_k when the objective function is quadratic, the proposed modified BFGS methods maintain the difficulty

associated with the BFGS method in this case. Due to this feature and the above observations, we consider only formula (6) here and state the following investigations.

In practice, we observed that formula (6) with σ_2 sufficiently close to 0.6 seems to work well in some cases, but generally worsens the performance of the D-BFGS method as σ_2 decreases. This observation illustrates that large values of σ_2 such as 0.8 or 0.9 are usually recommended when the D-BFGS update is employed in methods for constrained optimization, see for example (Fletcher, 1987; Nocedal and Wright, 1999; Powell, 1978, 2009). We will argue below that small values of φ_k are also useful in some cases, although large values of φ_k yield small changes in y_k.

Al-Baali (2011) extends the global convergence property of the restricted Broyden class of methods to that of D-Broyden, assuming that

$$(1-\nu_1)\frac{\overline{\Theta}_k}{\mu_k^2} \leq \Theta_k \leq \frac{1-\nu_2}{\mu_k}, \tag{9}$$

where $\nu_1, \nu_2 \in (0,1)$,

$$\overline{\Theta}_k = \frac{1}{1-b_k h_k}, \quad b_k = \frac{s_k^T B_k s_k}{s_k^T y_k}, \quad h_k = \frac{y_k^T B_k^{-1} y_k}{s_k^T y_k} \tag{10}$$

and $\mu_k \in (0,1]$ is given by

$$\mu_k = \frac{\varphi_k}{\varphi_k + (1-\varphi_k)b_k}. \tag{11}$$

We observe that $\overline{\Theta}_k < 0$, and that condition (9) hold for the choice $\Theta_k = 0$ (the BFGS option) and any values of ν_1 and ν_2. Thus, this condition adds another useful feature to the BFGS method over the other methods. Although prior to this result Θ_k was never chosen outside the interval $(\overline{\Theta}_k, 1]$, except for a well defined SR1 method (see for example Fletcher, 1987), condition (9) allows any real number Θ_k for sufficiently small values of μ_k (or φ_k).

Recently, for further restrictions on φ_k, Al-Baali (2011) extends the above superlinear convergence property to the damped methods and shows that

$$\varphi_k(\rho_k - 1) \to 0 \qquad \text{and} \qquad \mu_k^2(b_k h_k - 1) \to 0, \tag{12}$$

noting that $\rho_k = 1/b_k$.

Although this result suggests using a small value of $\varphi_k < 1$ when either term $\rho_k - 1$ or $b_k h_k - 1$ is sufficiently remotely away from zero, we use this value only when both terms satisfy the latter condition. This suggestion is due to the fact that the Broyden family of updates is reduced to a single update if $b_k h_k = 1$ (see for example Al-Baali, 1993). Therefore, since the first limit in (12) yields formula (6) (see Al-Baali, 1993), we consider employing it only when the value of the product $b_k h_k$ is sufficiently larger than one so that we obtain

$$\varphi_k = \begin{cases} \dfrac{\sigma_2}{1-\rho_k}, & \rho_k < \min(1-\sigma_2, \dfrac{h_k}{1+\sigma_4}) \\[2ex] \dfrac{\sigma_3}{\rho_k - 1}, & 1+\sigma_3 < \rho_k < \dfrac{h_k}{1+\sigma_4} \\[2ex] 1, & \text{otherwise}, \end{cases} \tag{13}$$

where $\sigma_4 \geq 0$. This modified formula ensures that the value of ρ_k or its modification lies in $[1-\sigma_2, 1+\sigma_3]$ when $b_k h_k > 1+\sigma_4$. If the value of $\sigma_4 = 0$ is used, then the damped-technique is always employed, except when the Broyden family is reduced to a single update. In particular, this value with $\sigma_2 = 0.8$ and $\sigma_3 = \infty$ reduces formula (13) to that of Powell (1978) for D-BFGS, except when $b_k h_k = 1$ and $\rho_k < 1-\sigma_2$. Indeed, for some choices of these parameters, we observed that formula (13) works better than (6).

We now consider the second limit of (12) which suggests using $\mu_k = \sigma_4 / (b_k h_k - 1)^{1/2}$ if $b_k h_k > 1+\sigma_4^2$, where $\sigma_4 \geq 0$, which can be substituted into (11) to obtain the corresponding value of φ_k. However, since σ_4 is chosen by the user and $\mu_k \leq \varphi_k$ if $\rho_k \leq 1$, we consider

$$\varphi_k = \begin{cases} \dfrac{\sigma_4}{\sqrt{b_k h_k - 1}}, & b_k h_k > 1+\sigma_4 \\ 1, & \text{otherwise}, \end{cases} \tag{14}$$

where, as above, $\sigma_4 \geq 0$. If $\sigma_4 = 0$ is used, then as for formula (13), the damped-technique is always employed, except when the Broyden family is reduced to a single update. Since $\mu_k \geq \mu_k^2$, Al-Baali (2011) considers formula (14) with the denominator replaced by $b_k h_k - 1$. Although other choices for φ_k exist, formulae (13) and (14) are considered here for illustration.

We note that the above formulae for φ_k are independent of the updating parameter Θ_k, so that the global convergence condition (9) may not hold for sufficiently large values of $|\Theta_k|$. Therefore, in this case, this condition can be enforced by choosing φ_k sufficiently small. For example, the equality in (9) could be solved for μ_k which can be substituted into (11) to obtain the largest acceptable value of φ_k. In this way, all the damped algorithms which we consider below for $\Theta_k \notin (\bar{\Theta}_k, 1)$ (the divergent Broyden options) solved the considered problem successfully. However, alternatively, replacing $b_k h_k - 1$ by

$$a_k = (b_k h_k - 1)\max(1, |\Theta_k|) \tag{15}$$

in formula (14), it follows that

$$\varphi_k = \begin{cases} \dfrac{\sigma_4}{\sqrt{a_k}}, & a_k > \sigma_4 \\ 1, & \text{otherwise}. \end{cases} \tag{16}$$

Using formula (16), we observed that all the damped algorithms which we consider below for several values of Θ_k from inside and outside the convex interval solved the problem successfully with performance significantly better than that of BFGS.

On the basis of the above discussions, we state the following outline for practical damped quasi-Newton methods.

We note that the strong Wolfe conditions in *Step 3* yield the Wolfe conditions (2)-(3). If, on every iteration, the choice $\hat{y}_k = y_k$ is used in *Step 5*, which follows from (1) and (13) with values of $\sigma_2 = 1$, $\sigma_3 = \infty$ and $\sigma_4 = \infty$, then Algorithm 2.1 is reduced to the 'undamped' Broyden family of methods. In particular, this choice with $\Theta_k = 0$, for all values of k, reduces the D-BFGS method to the standard BFGS method.

In practice, we observed that several values for $\sigma_2 > 1/2$ improve the performance of the BFGS and DFP methods (the latter significantly). If Θ_k is chosen such that the usual positive definiteness property holds (which we consider here), then any value of φ_k satisfies condition (9). Indeed, we observed that sufficiently small

values of φ_k are sometimes useful, particularly when $|\Theta_k|$ is remotely away from zero. These features will be illustrated in the next section.

Algorithm 2.1 D-Broyden class

Step 0: Given a starting point x_1, a symmetric and positive-definite matrix B_1, positive values of σ_0 and σ_1, and a tolerance $\varepsilon > 0$. Set $k = 1$.

Step 1: Terminate if $\| g_k \| \leq \varepsilon$.

Step 2: Compute the search direction $d_k = -B_k^{-1} g_k$.

Step 3: Find a steplength α_k and a new point $x_{k+1} = x_k + \alpha_k d_k$ such that the following strong Wolfe conditions hold:

$$f_k - f_{k+1} \geq -\sigma_0 \alpha_k d_k^T g_k, \quad | d_k^T g_{k+1} | \leq -\sigma_1 d_k^T g_k. \tag{17}$$

Step 4: Compute s_k, y_k, b_k and h_k.

Step 5: Choose values for Θ_k and φ_k.

Step 6: Update y_k to $\hat{y}_k$ and hence B_k to B_{k+1}, using (1) and (4), respectively.

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

3. A numerical example

The question that has been most useful to the development of successful algorithms for unconstrained optimization is 'Does the method work well when the objective function is quadratic?' This statement is given by Powell (2009) who also states that the answer is very welcome and encouraging for the updating of second derivative matrices of quadratic models by a Broyden family method. Since this family usually suffers from ill-conditioned problems, we would like to apply some selected Broyden family methods to a quadratic function with a very ill-conditioned Hessian matrix.

Because quasi-Newton methods are invariant under linear transformation, the difficulty is unchanged if we choose the simple quadratic function

$$f(x) = \frac{1}{2} x^T x, \quad x \in R^2$$

which has the solution $x^* = 0$ and identity Hessian $\nabla^2 f(x) = I$, and the initial Hessian approximation B_1 is very ill-conditioned. In particular, Powell (1986) uses

$$B_1 = \operatorname{diag}(1, \lambda),$$

where $\lambda > 0$, and

$$x_1 = \begin{pmatrix} \sqrt{c} \\ \sqrt{1-c} \end{pmatrix}, \quad c = \frac{1}{1+\lambda}.$$

The author simplifies the analytic derivation of x_{k+1} when calculated by either the BFGS or DFP method with $\alpha_k = 1$. He shows that for large values of λ, BFGS performs badly and DFP is far worse, although y_k is exact. Note that the former performance is maintained for the modified BFGS methods which use (8). The derivation of Powell (1986) for x_{k+1} has been extended to a restricted Broyden class of methods by Byrd *et al.* (1992) who also show that the performance of this class becomes worse as Θ_k increases. Since these papers illustrate that

these methods usually suffer from large eigenvalues of B_k rather than small ones, we will describe below the numerical results for a very large value of λ.

In a similar manner, we have tested the D-Broyden class of methods by applying Algorithm 2.1 to the above problem. To define *Step 0* of this algorithm, we let x_1 and B_1 be given as above, using $\lambda = 10^{10}$ and $\varepsilon = 10^{-7}$. We do not define σ_0 and σ_1 here, because $\alpha_k = 1$ was used in *Step 3* on all iterations whether this value satisfies the Wolfe conditions or not. Nevertheless, these conditions hold to the limit if a quasi-Newton method converges superlinearly (Dennis and Moré, 1974). Thus the algorithms we consider below differ only in the choices of Θ_k and φ_k as required in *Step 5* of Algorithm 2.1. All experiments were run in Matlab.

The damped-technique (1) with suitable values of φ_k has the ability of correcting the Hessian approximations successfully. Here, we report the results for several choices of φ_k and $\Theta_k \in \{-10^7, -100, -0.5, 0, 0.5, 1, 1.5, 100, 10^7\}$ which belong to the so called convex, preconvex and postconvex classes of methods (defined for $0 \le \Theta_k \le 1$, $\Theta_k < 0$ and $\Theta_k > 0$, respectively). We will report the number of function evaluations (nfe) which is required to terminate Algorithm 2.1 in *Step 1* with the Euclidean norm $\| g_k \| \le \varepsilon$. This number is the same as that of the gradient evaluations as well as that of iterations. 'F' indicates the algorithm was terminated before the latter inequality held, ie the algorithm failed to solve the problem. We also consider the BFGS method for comparison, which required $nfe = 32$ to solve the problem.

Table 1. D-BFGS with formula (6)

σ_2	0.95	0.9	0.7	0.6	0.5	0.4	0.1	0.01	0.001	10^{-6}
nfe	32	32	32	27	35	47	220	2107	18887	F

We firstly tested the D-BFGS method with φ_k given by formulae (6), (13) and (14) for various choices of the parameters σ_2 and σ_4. We do not consider choices for σ_3 here, because the second case in (6) and (13) was not used as the inequality $\rho_k > 1 + \sigma_3$ was not satisfied for $\sigma_3 > 0$. The results expressed in terms of nfe are presented in Table 1. We used (6) for 10 different values of $\sigma_2 \in (0,1)$. For sufficiently large values of σ_2, D-BFGS required $nfe = 32$ to solve the problem, which is the same as that required by BFGS. The reason is that using $\sigma_2 = 1$ in formula (6) yields $\varphi_k = 1$ which (by the damped-technique (1)) reduces the D-BFGS update to the BFGS one. We also note that D-BFGS works slightly better than BFGS when σ_2 is close to 0.6, while it becomes worse as σ_2 decreases in (0,0.5). Indeed, D-BFGS also failed to solve the problem when values of $\sigma_2 \le 10^{-6}$ were used. This observation is somehow expected on the basis of some results from the literature, because (according to our knowledge) values of $\sigma_2 = 0.8$ or 0.9 have been used in the D-BFGS method for constrained optimization during the last three decades, see for example (Powell, 2009).

This drawback of formula (6) with choices of $\sigma_2 \le 0.5$ can be avoided by employing it only when the BFGS update is sufficiently away from the rank one update, that is when $b_k h_k$ is sufficiently away from 1. To illustrate this feature, we repeated the run using formula (13) with most of the above values of σ_2 and some values of $\sigma_4 \ge 0$. The results are given in Table 2. We do not report the results for $\sigma_4 > 2$ here, because we observed that D-BFGS required the same nfe as that required for $\sigma_4 = 2$. It is clear that the damped-technique

improves BFGS substantially when σ_4 is defined sufficiently close to 0.5 and $\sigma_2 \le 0.6$ and significantly for very small values of σ_2. It is rather surprising that D-BFGS with $\sigma_2 = 10^{-6}$ and $\sigma_4 < 2$ required at most $nfe = 7$ to solve the problem. This result shows that small values of σ_2 (and hence φ_k) are useful in practice. This table also shows that the choice of $\sigma_4 = 0$ (or nearly so) which yields that the damped-technique is employed on most iterations, is not desirable.

Table 2. D-BFGS with formula (13)

$\sigma_4 \setminus \sigma_2$	0.95	0.9	0.7	0.6	0.5	0.4	0.1	0.01	10^{-3}	10^{-6}
2	32	32	32	32	32	32	32	32	32	32
1.5	32	32	32	20	18	17	12	8	7	6
0.95	32	32	32	20	18	17	12	8	8	6
0.5	32	32	32	20	18	17	12	8	8	5
0.1	32	32	32	20	19	18	12	8	8	5
0.001	32	32	32	22	20	19	13	8	8	5
10^{-6}	32	32	32	24	21	19	14	9	8	5
0	32	32	32	27	27	25	87	625	3918	7

The damped-technique seems to work well if the left limit in (12) is enforced only when the term under the right limit, $b_k h_k - 1$, is sufficiently away from zero. Therefore, it is worth testing formula (14) which enforces the right limit in (12). The results are given in Table 3. We observe that the damped-technique improves over BFGS as σ_4 decreases. For example when $\sigma_4 = 0.01$, D-BFGS required only $nfe = 8$, which is very small compared to 32 as required by BFGS.

Table 3. D-BFGS with formula (14)

σ_4	2	1	0.7	0.6	0.5	0.4	0.1	0.01	10^{-3}	10^{-6}	0
nfe	32	19	17	16	15	14	11	8	7	5	4

To see if the above encouraging numerical results for D-BFGS are generalised to other members of the D-Broyden class, we repeated the run for formula (14) with the same values of σ_4 and several values of Θ_k; some of them are very remotely away from 0 (the BFGS option). The numerical results are presented in Table 4.

In the first row of results, we used $\sigma_4 = \infty$ to obtain $\varphi_k = 1$ for all k and, hence, the corresponding undamped quasi-Newton methods were considered for comparison. The failures for $\Theta_k < 0$ and $\Theta_k > 1$ are expected, see for example (Byrd et al., 1992). Note that $nfe \approx \lambda = 10^{10}$ required for $\Theta_k = 1$ (the DFP option) is expected, see (Powell, 1986).

The remaining results in Table 4 were obtained by testing condition (9) with $v_1 = v_2 = 0.05$. If it does not hold for the computed φ_k, we reduce this value such that this condition holds with equality. In this way convergence is enforced to all algorithms which, indeed, solves the problem successfully. Thus the second row

of results with $\sigma_4 = \infty$ and condition (9) being enforced indicates that the damped-technique was employed for all algorithms defined with $\Theta_k \notin [0, 0.95]$.

Table 4. D-Broyden with formula (14) and enforcing condition (9)

$\sigma_4 \setminus \Theta_k$	-10^7	-100	-0.5	0	0.5	1	1.5	100	10^7
$\infty *$	F	F	F	32	78	$\approx 10^{10}$	F	F	F
∞	16	21	16	32	78	411	118	150	18424
2	17	9	8	32	78	411	118	65	18425
0.95	17	9	13	19	22	30	30	65	18425
0.5	14	10	15	15	16	18	21	65	18425
0.1	11	19	10	11	11	12	15	66	18425
0.01	11	7	8	8	8	10	13	65	18425
0.001	11	7	7	7	7	9	12	64	18425
10^{-6}	6	5	5	5	5	7	10	62	18425
0	7	7	7	7	7	9	12	79	22172
* Condition (9) was not enforced.									

An examination of the other results in Table 4 shows the following. The damped algorithms with $\Theta_k \leq 1.5$ and $\sigma_4 \leq 0.1$ work substantially better than BFGS. It is rather surprising that all algorithms with $\Theta_k < 0$, which sometimes yields indefinite Hessian approximations in the undamped algorithms, solved the problem successfully. Although BFGS works much better than the algorithms for $\Theta_k \geq 100$, it is also surprising that they solved the problem successfully. Note that the slow convergence of these algorithms is avoided as described in the following paragraph.

Table 5. D-Broyden with formula (16)

$\sigma_4 \setminus \Theta_k$	-10^7	-100	-0.5	0	0.5	1	1.5	100	10^7
$\infty *$	F	F	F	32	78	$\approx 10^{10}$	F	F	F
∞	16	21	16	32	78	411	118	150	18424
0.95	20	16	13	19	22	27	23	16	18
0.5	13	11	15	15	16	17	16	12	12
0.1	12	8	10	11	11	10	10	8	34
0.01	820	7	8	8	8	8	8	8	1109
0.001	7	24	7	7	7	7	7	10	6
10^{-6}	5	6	5	5	5	6	6	5	5
0	4	4	4	4	4	4	4	4	4
* Condition (9) was not enforced.									

We now repeat the test with formula (16). The results are given in Table 5. We observe the surprising results that all damped algorithms solve the problem successfully. For sufficiently small values of σ_4, the

algorithms performance is significantly better than that of BFGS, and the algorithms improve as σ_4 decreases. $\sigma_4 = 0$ means that the damped-technique is always employed, except when the Broyden family of updating formulae is reduced to the SR1 update. All algorithms approach an optimal method with $nfe = 4$ which is required to solve the problem by a well defined SR1 method. Although $\sigma_4 = 0$ gives the best choice, the value of $\sigma_4 = 0.95$ might be typical for a general function, and further experiments on several test problems should be considered.

It is worth reporting that an application of some algorithms to the above quadratic problem (or its linear transformation) with some choices for x_1 and $B_1 = \text{diag}(1, 10^{-10})$ has been considered. We observed, as expected, that all the algorithms solved the problem successfully with $nfe = 2$ or 3 whether the damped-technique was employed or not.

To see if the above results are generalised to general functions, we applied Algorithm 2.1, employing the strong Wolfe conditions (17) with the usual values of $\sigma_0 = 10^{-4}$ and $\sigma_1 = 0.9$, to some unconstrained optimization problems and observed the following. When Θ_k is sufficiently away from zero, the damped algorithms work significantly better than the undamped Broyden methods. For $\Theta_k = 0$, we observed that generally the standard BFGS method is preferable to the D-BFGS algorithm unless further restrictions on the damped parameter φ_k are made or, similarly, the scalars σ_2, σ_3 and σ_4 are appropriately chosen. Indeed, Al-Baali and Grandinetti (2009) showed that the performance of the D-BFGS method with formula (6) for $\sigma_2 = 0.9$, $\sigma_3 = 9$ and further restrictions on φ_k is substantially better than that of the BFGS method. This performance is improved further when both the damped and self-scaling techniques are combined in a certain sense (Al-Baali and Khalfan, 2009). For large-scale problems, Al-Baali (2004) introduced the damped formula (6) to the limited memory L-BFGS method, see for example (Nocedal and Wright, 1999). Using $\sigma_2 = 0.6$, $\sigma_3 \geq 3$ and sufficiently large values of φ_k, Al-Baali (2004) reported encouraging numerical results in certain cases.

4. Conclusion

It is shown that the D-Broyden class of methods with appropriate choices for the damped-technique work very well when applied to the quadratic function of Powell (1986). In particular, the damped parameter (16) improves the performance of robust methods (e.g. BFGS) substantially and inefficient methods (e.g. DFP) significantly and enforces fast convergence of divergent methods. The numerical results also demonstrate that small values for the damped parameter φ_k are useful in some cases.

Since this finding contradicts the well known fact that large values of $\varphi_k \geq 0.8$ should be used in the D-BFGS update of methods for the Lagrangian function in constrained optimization (see for example Fletcher, 1987; Nocedal and Wright, 1999; Powell, 1978; Powell, 2009; Gill and Leonard, 2003), it is worth noting that smaller values of φ_k using formulae (13) and (14) which modify that of Powell (1978) should also be further investigated.

For a general function, these formulae can be used with y_k replaced by the right hand side of (8) as in (Al-Baali and Grandinetti, 2009).

5. Acknowledgment

6. References

AL-BAALI, M. 1993. Variational quasi-Newton methods for unconstrained optimization. *JOTA*, **77**: 127-143.

AL-BAALI, M. 2003. Quasi-Wolfe conditions for Newton-like methods. *18th International Symposium on Mathematical Programming,* Copenhagen, Denmark, August 18-22, 2003.

AL-BAALI, M. 2004. Quasi-Wolfe conditions for quasi-Newton methods for large-scale optimization. *40th Workshop on Large-Scale Nonlinear Optimization,* Erice, Italy, June 22 - July 1, 2004.

AL-BAALI, M. 2011. Convergence analysis of a family of damped quasi-Newton methods for nonlinear optimization. *Research Report* DOMAS 11/2, Sultan Qaboos University, Oman.

AL-BAALI, M. and GRANDINETTI, L. 2009. On practical modifications of the quasi-Newton BFGS method. *Advanced Modeling and Optimization*, **11**: 63-76.

AL-BAALI, M. and KHALFAN, H.F. 2009. A combined class of self-scaling and modified quasi-Newton methods. *Optimization Online*, 07/2334 (to appear in *COAP*).

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.

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.

DENNIS, J.E. and SCHNABEL, R.B. 1996. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM Publications, Philadelphia.

FLETCHER, R. 1987. *Practical Methods of Optimization* (2nd edition). Wiley, Chichester, England.

GILL, P.E. and LEONARD, M.W. 2003. Limited-memory reduced-Hessian methods for large-scale unconstrained optimization. *SIAM J. Optim.*, **14**: 380-401.

GRATTON, S. and TOINT, P.L. 2010. Approximate invariant subspaces and quasi-Newton optimization methods. *OMS*, **25**: 507-529.

NOCEDAL, J. 1980. Updating quasi-Newton matrices with limited storage. *Math. Comput.*, **35**: 773-782.

NOCEDAL, J. and WRIGHT, S.J. 1999. *Numerical Optimization*. Springer, London.

POWELL, M.J.D. 1978. Algorithms for nonlinear constraints that use Lagrange functions. *Math. Programming*, **14**: 224-248.

POWELL, M.J.D. 1986. How bad are the BFGS and DFP methods when the objective function is quadratic?. *Math. Programming*, **34**: 34-47.

POWELL, M.J.D. 2009. On nonlinear optimization since 1959. *Optimization Online*, 11/2477.

YABE, H., OGASAWARA, H. and YOSHINO, M. 2007. Local and superlinear convergence of quasi-Newton methods based on modified secant conditions. *Journal of Computational and Applied Mathematics*, **205**: 617-632.