

# An Overview of Some Practical Quasi-Newton Methods for Unconstrained Optimization

M. Al-Baali\* and H. Khalfan\*\*

*Department of Mathematics and Statistics, College of Science, Sultan Qaboos University, P.O.Box 36, PC 123, Al-Khodh, Muscat, Sultanate of Oman, \*Email: albaali@squ.edu.om, \*\*Department of Mathematics Sciences, UAE University, Al-Ain 17551, United Arab Emirates, Email: humaidf@uaeu.ac.ae.*

دراسة حصرية لبعض الخوارزميات التطبيقية المشابهة لطريقة نيوتن للحلول المثلى غير المقيدة

محي الدين البعلي وحמיד خلفان

**خلاصة:** تعتبر الخوارزميات المشابهة لطريقة نيوتن من الطرق التطبيقية الأكثر كفاءة لمسائل الحلول المثلى غير المقيدة. تتضمن هذه المقالة دراسة شاملة لبعض الخوارزميات الحديثة من هذا النوع مع التركيز على الطرق المستخدمة في تقريب مصفوفة المشتقة الثانية ووسائل تطويرها.

**ABSTRACT:** Quasi-Newton methods are among the most practical and efficient iterative methods for solving unconstrained minimization problems. In this paper we give an overview of some of these methods with focus primarily on the Hessian approximation updates and modifications aimed at improving their performance.

**KEYWORDS:** Unconstrained optimization, line search technique, quasi-newton methods.

## 1. Introduction

In this paper we give an overview of some line search quasi-Newton methods for solving the unconstrained minimization problem

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

where  $f$  is a twice continuously differentiable function. Emphasis will be on the Hessian approximation formulas used in these methods, and techniques developed to improve their performance.

The basic iteration of a quasi-Newton method consists of the following. Starting with an initial approximation  $x_1$  to a solution  $x^*$  of (1), and an initial positive definite Hessian approximation  $B_1$ , calculate a new approximation at iteration  $k$  by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

where  $\alpha_k$  is a steplength and  $d_k$  is a search direction obtained by

$$d_k = -B_k^{-1} \nabla f(x_k), \tag{3}$$

where  $B_k$  is a positive definite  $n \times n$  Hessian approximation matrix chosen from the general Broyden class of updates discussed in the next section.

The steplength  $\alpha_k$  is calculated such that the Wolfe conditions

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \rho_1 \alpha_k \nabla f(x_k)^T d_k, \quad \rho_1 \in (0, 1/2) \tag{4}$$

and

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \rho_2 \nabla f(x_k)^T d_k, \quad \rho_2 \in (\rho_1, 1) \tag{5}$$

are satisfied. The first condition ensures sufficient reduction in  $f$ , and the second one guarantees that the steplength is not too small relative to the initial rate of decrease in  $f$ . In practice, the strong Wolfe conditions (4) and

$$|\nabla f(x_k + \alpha_k d_k)^T d_k| \leq -\rho_2 \nabla f(x_k)^T d_k$$

are usually used.

If beside the standard assumptions on  $f$ ,  $\nabla f$  is Lipschitz continuous, the steplength  $\alpha_k$  satisfies the Wolfe conditions, and the matrices  $B_k$  are positive definite and have a bounded condition number, then iteration (2) is globally convergent. (see e.g. Fletcher, 1987, Dennis and Schnabel, 1996, and Nocedal and Wright, 1999).

In the next section we discuss some well-known members of the Broyden class of Hessian approximation updates. In Section 3 we outline some approaches for improving the performance of some standard updates by modifications of the gradient difference  $\nabla f(x_{k+1}) - \nabla f(x_k)$ . In Section 4 we discuss self-scaling quasi-Newton methods aimed at handling ill-conditioned problems. Some quasi-Newton methods for large scale optimization are discussed in Section 5.

## 2. Quasi-newton updates

Setting  $\alpha_k = 1$  and  $B_k = \nabla^2 f(x_k)$  in (2) and (3), respectively, we obtain the well-known Newton method which converges quadratically to a solution  $x^*$ , if  $x_1$  is sufficiently close to  $x^*$ . For many practical problems, however, analytic second derivatives are unavailable, and the cost of approximating them by finite differences requires either an additional  $n$  gradient evaluations or  $O(n^2)$  function evaluations per iteration. Quasi-Newton updates for Hessian approximation avoid these disadvantages by approximating  $\nabla^2 f(x_{k+1})$  by a symmetric matrix

$$B_{k+1} = B_k + E_k,$$

where  $E_k$  is an  $n \times n$  matrix, such that the *secant equation*

$$B_{k+1} s_k = y_k, \tag{6}$$

is satisfied, where  $s_k = x_{k+1} - x_k$  and  $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ .

## AN OVERVIEW OF SOME PRACTICAL QUASI-NEWTON METHODS

The solution to equation (6), however, is not unique, and there is a variety of updating formulas obtained by imposing more conditions on  $B_{k+1}$ . If  $E_k$  is assumed to be a symmetric rank-one matrix for instance, then

$$(6) \text{ yields the unique Symmetric Rank-One update (SR1)} \quad B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{s_k^T y_k - s_k^T B_k s_k}. \quad (7)$$

Requiring  $E_k$  to be a symmetric rank-two matrix, equation (6) yields a variety of possible updates such as the well-known (broyden fletcher goldfarb shanno) and (davidon fletcher powell) formulas. A more general formula satisfying the secant equation with  $E_k$  symmetric and of rank two at most, known as the *Broyden class*, is given by

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

where

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

and  $\theta$  is a parameter. This class includes as special case the BFGS update, when  $\theta = 0$ ; the DFP update, when  $\theta = 1$ ; and the SR1 update, when  $\theta = s_k^T y_k / (s_k^T y_k - s_k^T B_k s_k)$ . (Another family of updates was proposed by Huang (1970), but is not discussed here since it was shown to be equivalent to the self-scaling Broyden family discussed in Section 5.)

If  $B_k$  is positive definite, and the *curvature condition*  $s_k^T y_k > 0$  holds (which is guaranteed by the second Wolfe condition (5)), then  $B_{k+1}$  is positive definite for any  $\theta > \bar{\theta}_k$ , where  $\bar{\theta}_k = \frac{y_k^T B_k s_k}{1 - b_k h_k}$ ,  $b_k = \frac{y_k^T B_k s_k}{s_k^T y_k}$ ,  $h_k = \frac{y_k^T H_k y_k}{s_k^T y_k}$

and  $H_k = B_k^{-1}$ . Since  $\bar{\theta}_k < 0$  (by Cauchy's inequality), it clearly follows that any update with  $\theta \geq 0$  (such as the BFGS and DFP updates) preserves positive definiteness if the curvature condition holds. The SR1 update preserves positive definiteness only if either  $b_k < 1$  or  $h_k < 1$ , which may not hold even for quadratic functions. (See e.g. Fletcher, 1987).

Powell (1976) showed that if  $f$  is a convex function and the Wolfe conditions hold, then for any starting point  $x_1$  and any positive definite initial matrix  $B_1$ , the BFGS method converges globally; and if furthermore the true Hessian  $\nabla^2 f(x^*)$  is positive definite, then the rate of convergence is  $q$ -superlinear. This result was extended by Byrd, Nocedal and Yuan (1987) to the interval  $0 \leq \theta < 1$ , (updates belonging to  $[0, 1]$  are known as the convex class).

Although Dixon (1972) showed that for a general nonlinear function  $f$ , all well-defined members of the Broyden family with  $\theta \neq \bar{\theta}_k$  generate the same sequence of iterates when used with exact line searches, numerical experience showed that only some updates, which we discuss next, worked well in practice when inexact line searches are used; and that the performance deteriorates as  $\theta$  increases above 0 (see e.g. Byrd, Liu and Nocedal, 1992).

The SR1 method enjoys some desirable features which are not shared by other standard updates. Fiaco and McCormick (1968) showed that for a positive definite quadratic function, if the SR1 update is used with linearly independent steps, and all the updates are well-defined, then the solution is reached in at most  $n+1$  iterations. Furthermore, if  $n+1$  iterations are required, then the final Hessian approximation  $B_{n+1}$  is the actual Hessian. This quadratic termination property is not generally true for other members of the Broyden family, unless exact line searches are used.

For general functions, Conn, Gould and Toint (1991) proved that the sequence of SR1 Hessian approximations converges to the true Hessian at the solution provided that the steps are uniformly linearly independent; that the SR1 update denominator is always sufficiently different from zero, and that the iterates converge to a finite limit. Hence under these conditions the rate of convergence is  $q$ -superlinear. If the assumption of uniform linear independence is dropped, then as shown in Khalfan, Byrd and Schnabel (1993) the SR1 method converges  $(n+1)$ - $q$ -superlinearly provided that for all  $k$ ,  $B_k$  is positive definite and bounded.

On the other hand, Ge and Powell (1983) showed that the sequence of matrices generated by the BFGS method converges to a matrix not necessarily equal to the true Hessian  $h_k \leq 2b_k h_k$ .

In order to obtain a well-conditioned update, Davidon (1975) proposed the update

$$\theta = \begin{cases} \frac{1}{1-b_k}, & \text{otherwise.} \end{cases} \tag{8}$$

The first update in this formula is obtained by minimizing over  $\theta$  the condition number  $\kappa(B_{k+1}(\theta)H_k)$ , and the second one is the SR1 formula. Practical experience, however, showed no significant improvement using this update (see e.g. Al-Baali, 1993 and Lukšan and Spedicato, 2000).

Since updates from the preconvex class work well in practice (see e.g. Zhang and Tewarson 1988 and Byrd, Liu and Nocedal, 1992), Al-Baali (1993) reported improved numerical performance using the switching BFGS/SR1 update

$$\theta = \begin{cases} \frac{1}{1-b_k}, & \text{otherwise,} \end{cases}$$

which preserves positive definiteness. Lukšan and Spedicato (2000) also reported competitive performance using this update. Other updates of the switching type are given in Al-Baali, Fuduli and Musmanno (2004).

An interval of globally convergent updates was tested by Al-Baali and Khalfan (2005), defined by  $\theta_k^- \leq \theta \leq \theta_k^+$ ,

where

$$\theta_k^\pm = 1 \pm \frac{1}{v_k}, \quad v_k = \left(1 - \frac{1}{b_k h_k}\right)^{1/2}. \tag{9}$$

Their numerical experience indicated that several updates from this interval worked well in practice, especially the modified SR1 update

$$\theta = \begin{cases} \max\left(\theta_k^-, \frac{1}{1-b_k}\right), & b_k > 1 \\ \min\left(\theta_k^+, \frac{1}{1-b_k}\right), & b_k < 1 \end{cases}$$

and the update defined by

$$\theta = \max(\theta_k^-, 1 - b_k)$$

which does not include the SR1 formula. The fast rate of convergence observed in the numerical experience with these updates suggests further study of their convergence properties.

### 3. Modifying gradient-difference vector

Many approaches have been proposed to improve the quasi-Newton Hessian approximation updates. In this section we outline some recent suggested updates obtained by modifying the vector  $y_k$ .

Zhang, Deng, and Chen (1999) suggested replacing  $y_k$  in the BFGS formula by the vector

$$y_k^* = y_k + \frac{(g_k - f_{k+1}) + (g_{k+1} + g_k) s_k}{\|s_k\|^2} s_k \tag{10}$$

which has the property that

$$\frac{s_k^T}{\|s_k\|} [\nabla^2 f(x_{k+1}) s_k - y_k^*] = O(\|s_k\|^3)$$

and reduces to  $y_k$  if  $f$  is quadratic. The resulting modified BFGS method retains global and  $q$ -superlinear convergence for convex functions, and performs slightly better than the standard BFGS method on some test problems.

Li and Fukushima (2001) proposed using

$$y_k = y_k + \left[ \|g_k\| + \max\left(\frac{-y_k^T s_k}{\|s_k\|^2}, 0\right) \right] s_k$$

instead of  $y_k$  in the BFGS update for ensuring that  $y_k^{*T} s_k > 0$ . They showed that for a general function, the resulting BFGS method with backtracking line search converges globally and superlinearly, under standard assumptions on the objective function. Their numerical experience also indicated some improvement in performance. For other modifications of this type (Yuan, 1991, Xu and Zhang, 2001, Wei *et al.*, 2004, and Zhang, 2005).

Modifying  $y_k$  was originally suggested by Powell (1978) who proposed a BFGS method for constrained optimization with

$$y_k^* = y_k + (1 - \varphi)(B_k s_k - y_k),$$

where  $0 < \varphi \leq 1$ . Al-Baali (2003b) used this modification for unconstrained problems and reported improved performance for the BFGS and many other methods.

Other approaches for improving Hessian approximation involved employing multiple quasi-Newton updates at each iteration, using information at the current and previous steps. Some improvement was observed for certain type of test problems (Khalfan, 1989, Ford and Tharmlikit, 2003, and Al-Baali, *et al.*, 2004).

#### 4. Self-scaling quasi-newton methods

The standard quasi-Newton methods that we considered so far may have difficulties in solving some ill-conditioned problems. Powell (1986) showed that, the BFGS and DFP methods behaved badly (the latter far worse) when applied to a simple ill-conditioned quadratic function. Moreover, Dai (2002) and Mascarenhas (2004) gave examples of nonconvex functions, for which the BFGS method failed to converge to the solution of the problem. In this section we consider self-scaling Hessian approximation updates for handling ill-conditioned problems.

Oren and Luenberger (1974) proposed the two parameters class of self-scaling Hessian approximation updates,

$$B_{k+1}(\theta, \tau) = \tau \left( B_k - \frac{B_k S_k S_k^T B_k}{S_k^T B_k S_k} \right) + \frac{y_k y_k^T}{y_k^T S_k} + \tau \theta w_k w_k^T, \tag{11}$$

where  $\theta$  and  $\tau$  are chosen such that the new update is optimally conditioned in some sense. For the parameters  $\theta$  and  $\tau$  the authors suggested the intervals  $\theta \leq 1, \frac{1}{b_k} \leq \tau \leq h_k$

to ensure that  $\kappa(B_{k+1}(\theta, \tau)G^{-1}) \leq \kappa(B_k G^{-1})$  when  $f$  is a quadratic function with a positive definite Hessian  $G$ .

Using intervals (12), Oren and Spedicato (1976) proposed the class

$$\tau (b_k h_k - 1) \tag{13}$$

which can be obtained by minimizing the condition number  $\kappa(B_{k+1}(\theta, \tau)H_k / \tau)$  over  $\theta$  (see Spedicato (1978)). Notice that substituting  $\tau = 1$  in class (13) gives the Davidon optimally conditioned update, given in the first case of (8). Moreover, as shown in Al-Baali (1995), class (13) can be also obtained by minimizing the same condition number over  $\tau$ .

In order to include members from the nonconvex class such as the SR1 update, and maintain positive definiteness, Al-Baali (1995) and Hu and Storey (1994) proposed the intervals

$$\theta_k^- \leq \theta \leq \theta_k^+, \tau_k^- \leq \tau \leq \tau_k^+, \tag{14}$$

where  $\tau_k^\pm = h_k (1 \pm v_k)$ , and where  $\theta_k^\pm$  and  $v_k$  are defined by (9). The end points of intervals (14) define two self-scaling SR1 updates, which are the same updates obtained by Osborne and Sun (1988) and Wolkowicz (1996). The choice  $\theta = 1 / (1 - b_k \tau_k^+) \leq 0$  with  $\tau = \tau_k^+$  is preferable since methods from the preconvex class work well in practice.

## AN OVERVIEW OF SOME PRACTICAL QUASI-NEWTON METHODS

Most of these self-scaling updates, however, do not generally improve the performance of the unscaled methods as reported by several authors. In fact, Shanno and Phua (1978) showed that the BFGS method worked better when scaling is used only for the initial Hessian approximation  $B_1$ . Moreover, Nocedal and Yuan (1993) showed that, compared with the (unscaled) BFGS method, the best self-scaling BFGS algorithm of Oren and Luenberger (1974),  $\theta = 0$  and  $\tau = 1/b_k$ , performs badly, when used for solving a simple quadratic problem of two variables. They also showed that for the same problem, superlinear convergence is not obtained unless certain step length values are used which cannot be guaranteed in practice.

Al-Baali (1998) however, extended the global and superlinear convergence theory of Byrd, Liu and Nocedal (1992) for convex functions to the self-scaling class, (11), on the intervals

$$c_1 \leq \tau < 1, \quad c_2 \theta_k \leq \tau \theta \leq c_3, \quad (15)$$

where  $c_1, c_2, c_3 \in (0, 1)$ , and reported that self-scaling methods from these intervals, outperformed corresponding unscaled methods. Further numerical testing reported by Al-Baali and Khalfan (2005) showed that these methods succeeded in solving more problems than the unscaled methods, especially when  $\theta \geq 1$ .

Using scaling only when  $\tau < 1$ , also improves the performance of other self-scaling methods discussed above. For example, replacing the best self-scaling BFGS method of Oren and Luenberger (1974) mentioned above with

$$\tau = \min \left( \frac{1}{b_k}, 1 \right),$$

the resulting method outperformed even the standard BFGS method. Performance improvement was also reported, especially for the DFP method, by Contreras and Tapia (1993), and Yabe, Martines, and Tapia (2004), when a similar self-scaling approach was used for a certain type of problems.

### 5. Large-scale quasi-newton methods

The storage and computational requirement of the methods we considered so far is  $O(n^2)$  for a problem of  $n$  variables. Several modifications of quasi-Newton updates have been proposed to improve their efficiency when this cost is excessive. In this section we consider two approaches that are used widely in practice: the limited-memory method which is suitable for problems in which the true Hessian is not sparse, and the partitioned method which is the method of choice for problems with partially separable Hessians.

#### 5.1 Limited-memory methods

In these methods only a few vectors of length  $n$  are used for approximating the inverse of the Hessian implicitly instead of storing a full  $n \times n$  matrix. For example, if we write the BFGS update in the form

$$H_{k+1} = V_k^T H_k V_k + \rho_k I, \quad V_k = I - \rho_k y_k s_k^T, \quad \rho_k = 1/s_k^T y_k,$$

then using the  $m$  vector pairs  $\{s_i, y_i\}, i = k, k-1, \dots, k-m+1, H_{k+1}$  can be expressed as

$$\begin{aligned}
H_{k+1} = & \left( V_k^T \cdots V_{k-m+1}^T \right) H_{k-m+1} \left( V_{k-m+1} \cdots V_k \right) \\
& + \rho_{k-m+1} \left( V_k^T \cdots V_{k-m+2}^T \right) S_{k-m+1} S_{k-m+1}^T \left( V_{k-m+2} \cdots V_k \right) \\
& + \rho_{k-m+2} \left( V_k^T \cdots V_{k-m+2}^T \right) S_{k-m+1} S_{k-m+1}^T \left( V_{k-m+2} \cdots V_k \right) \\
& + \cdots \\
& + \rho_k S_k S_k^T.
\end{aligned}$$

In the limited-memory method of Nocedal (1980), the matrix  $H_{k-m+1}$  is replaced at each iteration by a positive definite diagonal matrix  $D_k$ , and the search direction  $d_{k+1} = -H_{k+1} \nabla f(x_{k+1})$  is obtained by  $4mn$  multiplications involving the  $m$  pairs  $\{s_i, y_i\}$  and  $\nabla f(x_{k+1})$  as shown for example in Nocedal and Wright (1999). This method is referred to as the L-BFGS method and it converges only globally for convex functions as shown in Liu and Nocedal (1989). A compact representation of limited-memory methods for general quasi-Newton updates is given by Byrd, Nocedal and Schnabel (1994).

Computational experience with  $D_k = (1/h_k)I$ , a multiple of the identity matrix, indicates that for large scale problems in which  $\nabla^2 f(x_{k+1})$  is not sparse, the L-BFGS method outperforms other methods such as the nonlinear conjugate gradient method; and that its performance improves substantially, in term of computing time, as  $n$  gets large. The method however may suffer from slow convergence, which costs more function evaluations, especially on very ill-conditioned problems (Nocedal and Wright, 1999). For large-scale least-square problems, Al-Baali (2003a) considered a modified L-BFGS method using a vector  $y_k^*$ , similar to the one discussed in section 3, instead of  $y_k$  and reported substantial improvement in numerical performance.

Another limited-memory approach, is based on the fact that the standard BFGS method accumulates approximate curvature in a sequence of expanding subspaces, which allows using a smaller reduced matrix to approximate the Hessian, that increases in dimension at each iteration. This feature is used to define limited-memory reduced-Hessian methods that require half the storage of conventional limited-memory methods. For more on these methods see Gill, and Leonard (2003 ) and Lukšan and Vlček (2006).

## 5.2 Partitioned methods

Every function  $f$  with a sparse Hessian can be written in the form

$$f(x) = f_1(x) + f_2(x) + \cdots + f_m(x),$$

where each function  $f_i$  depends only on a few variables  $n_i$ , and  $n = n_1 + n_2 + \cdots + n_m$  (see Griewank and Toint (1982)). In a partitioned method the Hessian of each element function  $f_i$  is approximated using a quasi-Newton Hessian approximation,  $B_k^i$ . These matrices are then assembled to define a sparse Hessian approximation  $B_k$  to  $\nabla^2 f(x_k)$  for finding the search direction by solving

$$B_k d_k = -\nabla f(x_k). \quad (16)$$

If all the element functions  $f_i(x)$  are convex, then the BFGS method converges globally, even if the system (16 ) is solved inexactly as shown in Griewank (1991). The partitioned BFGS method performs well in practice provided that partial separability is fully exploited. Practical implementation however, mostly use the

## AN OVERVIEW OF SOME PRACTICAL QUASI-NEWTON METHODS

SR1 method, since some element functions may have indefinite Hessians (see Griewank and Toint (1984), Conn, Gould and Toint (1992, 1996).

Finally we mention that there are many other modifications of quasi-Newton methods that we did not cover in this paper. For other reviews of quasi-Newton methods see for instance Nocedal (1992) and Lukšan and Spedicato (2000).

### 6. References

- AL-BAALI, M. 1993. Variational quasi-newton methods for unconstrained optimization. *JOTA*, **77**: 127–143.
- AL-BAALI, M. 1995. On measure functions for the self-scaling updating formulae for quasi-newton methods. *Optimization*, **32**: 59–69.
- AL-BAALI, M. 1998. Global and superlinear convergence of a restricted class of self-scaling methods with inexact line searches for convex functions. *COAP*, **9**: 191–203.
- AL-BAALI, M. 2003a. Quasi-Newton Algorithms for Large-Scale Nonlinear Least-Squares. In High Performance Algorithms and Software for Nonlinear Optimization, Editors G. Di Pillo and A. Murli, Kluwer Academic, pp. 1-21.
- AL-BAALI, M. 2003b. *18th International Symposium on Mathematical Programming*, Copenhagen, August 18–22.
- AL-BAALI, M., FUDULI, A. and MUSMANNO, R. 2004. On the Performance of Switching BFGS/SR1 Algorithms for Unconstrained Optimization. *OMS*, **19**: 153-164.
- AL-BAALI, M. and KHALFAN, H. 2005. A Wide Interval for Efficient Self-Scaling Quasi-Newton Algorithms. *OMS*, **20**: 679–691.
- 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.
- BYRD, R.H., NOCEDAL, J. and SCHNABEL, R.B. 1994. Representation Matrices and their Use in Limited-Memory Methods. *Math. Programming*, Series A, **63**: 129-156.
- 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.
- CONN, A.R., GOULD, N.I.M. and TOINT, P.H.L. 1991. Convergence of Quasi-Newton Matrices Generated by the Symmetric Rank One Update. *Math. Programming*, **50**: 177–195.
- CONN, A.R., GOULD, N.I.M. and TOINT, P.H.L. 1991. LANCELOT: *A Fortran Package for Large-Scale Nonlinear Optimization* (Release A), Springer Series in Computational Mathematics, Springer, New York.
- CONN, A.R., GOULD, N.I.M. and TOINT, P.H.L. 1994. *Large-Scale Nonlinear Constrained Optimization: A Current Survey*, in *Algorithms for Continuous Optimization: The State of the Art* (E. Spedicato, ed.), Vol. 434 of NATO ASI Series C: Mathematical and Physical Sciences, Kluwer, Dordrecht, Netherlands, pp. 287–332.
- CONN, A.R., GOULD, N.I.M. and TOINT, P.H.L. 1996. Large-Scale Nonlinear Optimization 351, ‘Numerical Experiments with the LANCELOT Package (Release A) for Large-Scale Nonlinear Optimization.’ *Math. Programming*, Ser. A, **73**: 73–110.
- CONTRERAS, M. and TAPIA, R.A. 1993. Sizing the BFGS and DFP Updates: A Numerical Study. *JOTA*, **78**: 93–108.
- DAI, Y. 2002. Convergence Properties of the BFGS Algorithm. *SIAM J. Optim.*, **13**: 693–701.
- DAVIDON, W.C. 1975. Optimally Conditioned Optimization Algorithms without Line Searches. *Math. Programming*, **9**: 1-30.
- DENNIS, J.E. and SCHNABEL, R.B. 1996. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM Publications.
- DIXON, L.C.W. 1972. Quasi-Newton Methods Generate Identical Points. *Math. Programming*, **2**: 383–387.
- FIACO, A.V. and MCCORMICK, G.P. 1968. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York. Reprinted by SIAM Publications, 1990.

- FLETCHER, R. 1987. *Practical Methods of Optimization* (2nd edition), Wiley, Chichester, England. Reprinted in 2000.
- FLETCHER, R. 1994. An Overview of Unconstrained Optimization. In E. Spedicato, editor, *Algorithms for Continuous Optimization: The state of the Art*, Kluwer Academic, Boston, pp. 109–143.
- FORD, J.A. and THARMLIKIT, S. 2003. New Implicit Updates in Multi-Step Quasi-Newton Methods for Unconstrained Optimization. *J. Compt. Appl. Math.*, **152**: 133–146.
- GE, R.-P. and POWELL, M.J.D. 1983. The Convergence of Variable Metric Matrices in Unconstrained Optimization. *Math. Programming*, **27**: 123–143.
- GILL, P.E. and LEONARD, M.W. 2003. Limited-Memory Reduced-Hessian Methods for Large-Scale Unconstrained Optimization. *SIAM J. Optim.*, **14**: 380–401.
- GRIEWANK, A. 1991. The Global Convergence of Partitioned BFGS on Problems with Convex Decomposition and Lipschitzian Gradient. *Math. Programming*, **50**: 141-175.
- GRIEWANK, A. and TOINT, P.H.L. 1982. *On the Unconstrained Optimization of Partially Separable Objective Functions in Nonlinear Optimization*. In 1981 (M.J.D. Powell Ed.) Academic Press, (London), 301-302.
- GRIEWANK, A. and TOINT, P.H.L. 1984. *Numerical Experiments with Partially Separable Optimization Problems*. In *Numerical Analysis Proceedings Dundee 1983*, Lecture Notes in Mathematics 1066 (D.F. Griffiths Ed.) Springer Verlag (Berlin), 203-220.
- KHALFAN, H.F. 1989. Topics in Quasi-Newton Methods for Unconstrained Optimization, Ph.D. thesis, University of Colorado at Boulder, USA.
- KHALFAN, H.F., BYRD, R.H. and SCHNABEL, R.B. 1993. A Theoretical and Experimental Study of the Symmetric Rank One Update. *SIAM J. Optim.*, **3**: 1–24.
- HU, Y.F. and STOREY, C. 1994. Family of Optimally Conditioned Quasi-Newton Updates for Unconstrained Optimization. *JOTA*, **83**: 421–431.
- HUANG, H.Y. 1970. Unified Approach to Quadratically Convergent Algorithms for Function Minimization. *JOTA*, **5**: 405–423.
- LI, D. and FUKUSHIMA, M. 2001. A Modified BFGS Method and its Global Convergence in Nonconvex Minimization. *J. Compt. Appl. Math.*, **129**: 15–35.
- LIU, D.C. and NOCEDAL, J. 1989. On the Limited Memory BFGS Method for Large Scale Optimization. *Math. Programming (Series B)*, **45**:503–528.
- LUKŠAN, L. and SPEDICATO, E. 2000. Variable Metric Methods for Unconstrained Optimization and Nonlinear Least Squares. *J. Compt. Appl. Math.*, **124**: 61–95.
- LUKŠAN, L. and VLČEK, J. 2006. Variable Metric Method for Minimization of Partially Separable Nonsmooth Functions. *Pacific Journal of Optimization*, **2**: 59-70.
- MASCARENHAS, W.F. 2004. The BFGS Method with Exact Line Searches Fails for Nonconvex Objective Functions. *Math. Programming*, **99**: 49–61.
- NOCEDAL, J. 1980. Updating Quasi-Newton Matrices with Limited Storage. *Math. Comput.*, **35**: 773–782.
- NOCEDAL, J. 1992. Theory of Algorithms for Unconstrained Optimization. *Acta Numerica*, **1**: 199–242.
- NOCEDAL, J. and WRIGHT, S.J. 1999. *Numerical optimization*, Springer, London.
- NOCEDAL, J. and YUAN, Y. 1993. Analysis of a Self-Scaling Quasi-Newton Method. *Math. Programming*, **61**: 19–37.
- OREN, S.S. and LUENBERGER, D.G. 1974. Self-Scaling Variable Metric (SSVM) Algorithms, Part I: Criteria and Sufficient Conditions for Scaling a Class of Algorithms. *Manage. Sci.*, **20**: 845–862.
- OREN, S.S. and SPEDICATO, E. 1976. Optimal Conditioning of Self-Scaling Variable Metric Algorithms. *Math. Programming*, **10**: 70–90.
- OSBORNE, M.R. and SUN, L.P. 1988. A New Approach to the Symmetric Rank-One Updating Algorithm. Tech. Report, NMO/01, School of Mathematics, Australian National University.
- POWELL, M.J.D. 1976. Some Global Convergence Properties of a Variable Metric Algorithm for Minimization without Exact Line Searches. In R.W. Cottle and C.E. Lemke, editors, *Nonlinear Programming*. SIAM-AMS Proceedings, Vol. IX, SIAM Publications.

## AN OVERVIEW OF SOME PRACTICAL QUASI-NEWTON METHODS

- 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.
- SHANNO, D.F. and PHUA, K.H. 1978. Matrix Conditioning and Nonlinear Optimization. *Math. Programming*, **14**: 149–160.
- Spedicato, E. 1978. On a Conjecture of Dixon and Other Topics in Variable Metric Methods. *Math. Programming*, **15**: 123–129.
- WEI, Z., YU, G., YUAN, G. and LIAN, Z. 2004. The Superlinear Convergence of a Modified BFGS-Type Method for Unconstrained Optimization. *COAP*, **29**: 315–332.
- WOLKOWICZ, H. 1996. An Efficient Region of Optimal Updates for Least Change Secant Methods. *SIAM J. Optim.*, **5**: 172–191.
- XU, C. and ZHANG, J. 2001. A Survey of Quasi-Newton Equations and Quasi-Newton Methods for Optimization. *Annals of Operations research*, **103**: 213–234.
- YABE, H., MARTINEZ, H.J. and TAPIA, R.A. 2004. On Sizing and Shifting the BFGS Update within the Sized-Broyden Family of Secant Updates. *SIAM J. Optim.*, **15**: 139–160.
- YUAN, Y. 1991. A Modified BFGS Algorithm for Unconstrained Optimization. *IMA J. Numerical Analysis*, **11**: 325–332.
- ZHANG, J.Z., DENG, N.Y. and CHEN, L.H. 1999. Quasi-Newton Equation and Related Methods for Unconstrained Optimization. *JOTA*, **102**: 147–167.
- ZHANG, L. 2005. A Globally Convergent BFGS Method for Nonconvex Minimization without Line Searches. *OMS*, **20**: 737–747.
- 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.
- YUETINGA, Y. and XU, C. 2007. A Compact Limited Memory Method for Large Scale Unconstrained Optimization. *European Journal of Operational Research*, **180**: 48–56.
- 

Received 13 February 2007

Accepted 9 October 2007