Applications of Mathematics

Mahsa Nosrati; Keyvan Amini

A new diagonal quasi-Newton algorithm for unconstrained optimization problems

Applications of Mathematics, Vol. 69 (2024), No. 4, 501–512

Persistent URL: http://dml.cz/dmlcz/152531

Terms of use:

© Institute of Mathematics AS CR, 2024

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* http://dml.cz

A NEW DIAGONAL QUASI-NEWTON ALGORITHM FOR UNCONSTRAINED OPTIMIZATION PROBLEMS

Mahsa Nosrati, Keyvan Amini, Kermanshah

Received February 27, 2024. Published online July 15, 2024.

Abstract. We present a new diagonal quasi-Newton method for solving unconstrained optimization problems based on the weak secant equation. To control the diagonal elements, the new method uses new criteria to generate the Hessian approximation. We establish the global convergence of the proposed method with the Armijo line search. Numerical results on a collection of standard test problems demonstrate the superiority of the proposed method over several existing diagonal methods.

Keywords: unconstrained optimization; diagonal quasi-Newton method; weak secant equation; global convergence

MSC 2020: 65K05, 90C30

1. Introduction

In this study, the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) \tag{1}$$

is considered, where $f\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth function. A basic iterative process to solve (1) is

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

in which $d_k \in \mathbb{R}^n$ is a search direction, t_k is a positive step length, and x_0 is a given initial point. The success of the process (2) depends on effective choices of both the direction d_k and the step length t_k. In this regard, the step length t_k is selected so that it can cause a sufficient reduction in the function value of f and the direction d_k fulfills the descent condition, i.e., $\nabla f(x_k)^\top d_k < 0$.

It is known that choosing an appropriate line search is necessary to have an effective algorithm. There are various line search algorithms that can be referred to line

DOI: 10.21136/AM.2024.0045-24

© Institute of Mathematics, Czech Academy of Sciences 2024.

searches, such as Armijo [6], Wolfe [24] or Goldstein [16]. Among them, the Armijo line search is a popular method that needs the step length to fulfill the condition

$$f(x_k + t_k d_k) \leqslant f(x_k) + \sigma t_k g_k^\top d_k, \tag{3}$$

in which $\sigma \in (0,1)$ and $g_k = \nabla f(x_k)$. Choosing the descent direction d_k is another critical factor in the line search algorithms. Various processes for generating d_k have been suggested, such as gradient, Newton and quasi-Newton, conjugate gradient and trust region methods, see [21]. Among these processes, the quasi-Newton methods are effective iterative ones to solve the problem (1) due to the strong convergence properties (see [10], [11]). The customary iteration of this family is

$$x_{k+1} = x_k - t_k B_k^{-1} g_k,$$

where the matrix B_k is a symmetric approximation of the Hessian matrix $\nabla^2 f(x_k)$. For the next iteration, the updated Hessian approximation B_{k+1} should fulfill the secant equation

$$\begin{gathered} B_{k+1} s_k = y_k, \\ s_k = x_{k+1} - x_k, \quad y_k = g_{k+1} - g_k. \end{gathered} \tag{4}$$

Although it is well known that these methods have good numerical and theoretical results, there are some major defects in these methods. They require full matrix storage to approximate the Hessian matrix or its inverse. Thus, they are not appropriate for solving large-scale optimization problems. Some modified versions of quasi-Newton algorithms, such as truncated Newton methods [19] and limited-memory quasi-Newton methods [20] are designed to solve a large-scale problem while implementing these strategies is very complicated.

To create some easy minimization algorithms for unconstrained optimization, Gill and Murray [15] proposed that the Hessian matrix of the minimizing function can be approximated by a diagonal matrix with positive diagonal elements. They only retained the diagonal elements of the BFGS updating matrix. It is obvious that a diagonal matrix does not satisfy the secant equation (4). Based on this fact, Dennis and Wolkowicz [12] suggested a weak secant equation as

$$s_k^\top B_{k+1} s_k = s_k^\top y_k. \tag{5}$$

There are many various ideas for computing B_{k+1} satisfying (5). One of the techniques is based on some variational property similar to the technique used in PSB

updates [22], [9]. In this approach, the diagonal matrix B_{k+1} satisfying the weak secant equation (5) is determined by solving the subproblem

$$\min_{B\in\mathfrak{D}} \|B - B_k\| \quad \text{subject to} \quad s_k^\top B s_k = s_k^\top y_k, \tag{6}$$

where $\|\cdot\|$ is an arbitrary norm defined on $\mathbb{R}^{n\times n}$ and

$$\mathfrak{D} = \{B \in \mathbb{R}^{n\times n};\ B \text{ is diagonal}\}.$$

It is clear that selecting different matrix norms in (6) can generate some different diagonal quasi-Newton methods. Using the Frobenius norm, Zhu et al. [25] got the update formula

$$B_{k+1} = B_k + \frac{s_k^\top y_k - s_k^\top B_k s_k}{s_k^\top E_k^2 s_k} E_k^2, \tag{7}$$

where

$$E_k = \text{diag}(s_k^{(1)}, s_k^{(2)}, \ldots, s_k^{(n)}). \tag{8}$$

They claimed that the diagonal quasi-Newton (DQN) algorithm based on (7) has some appropriate numerical performances. They also showed that this method has linear convergence. In 2021, to improve the above update formula, Leong et al. [17] by using the weighted Frobenius norm

$$\|A\|_W^2 = \text{tr}(W^{-1}AW^{-1}A^\top)$$

proposed the update formula

$$B_{k+1} = B_k + \frac{s_k^\top y_k - s_k^\top B_k s_k}{s_k^\top (W E_k^2 W) s_k} W E_k^2 W, \tag{9}$$

where the operator $\text{tr}(M)$ defines the trace of the matrix M, W is a positive definite matrix, and E_k is defined by (8). Clearly, if W is chosen to be an $n \times n$ identity matrix I_n, the formula (9) reduces to (7). In addition, they found that $W = B_k$ is one of the best choices. In 2019, Andrei [4] considered the approximation matrix B_{k+1} as

$$B_{k+1} = B_k + \Delta_k,$$

in which the diagonal matrix Δ_k can be computed by solving the constrained optimization subproblem

$$\min_{\Delta_k\in\mathfrak{D}} \frac{1}{2}\|\Delta_k\|_F^2 + \text{tr}(B_k + \Delta_k) \quad \text{subject to} \quad s_k^\top (B_k + \Delta_k) s_k = s_k^\top y_k, \tag{10}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. To control the eigenvalues of the updated matrix B_{k+1}, the second term in the objective function of (10), $\mathrm{tr}(B_k+\Delta_k)$, is added. By solving the subproblem (10), Andrei introduced the direction

$$d_{k+1} = -B_{k+1}^{-1} g_{k+1},$$

where B_{k+1} is given by

$$B_{k+1} = B_k + \frac{s_k^\top y_k + s_k^\top s_k - s_k^\top B_k s_k}{s_k^\top E_k^2 s_k} E_k^2 - I_n, \tag{11}$$

and E_k is defined by (8). It is clear that the positive definiteness of this update is not guaranteed. Thus, the search direction d_{k+1} may not be a descent direction. To cope with this shortcoming, Andrei modified the direction d_{k+1} by enforcing some simple technique. He also claimed that this algorithm is well defined and converges q-linearly to a solution under Wolfe line search. Numerical experiments showed the numerical efficiency of this algorithm compared to some other diagonal methods. It is noticeable that the DQN direction can be considered as a spectral gradient direction; see, e.g., [23], [7], [18], [14].

Applying the same pattern of the DQN method, we get another approach introduced by Andrei [3] in which the diagonal elements are determined by minimizing the measure of the Byrd and Nocedal function. In 2021, an accelerated DQN algorithm was proposed by Andrei [5]. In his approach, he used scaling of the forward finite differences directional derivative of the gradient components as the elements of the diagonal matrix approximating the Hessian matrix.

In this paper, we present a new DQN method by modifying the subproblem (10). Under Armijo line search, the global convergence of the new algorithm is established. Numerical results show that the proposed method is more promising.

The rest of the paper is as follows. In the next section, we explain the main idea of the new method. In Section 3, the convergence analysis of the proposed algorithm is investigated. Numerical results are presented in Section 4. Some conclusions are delivered in the last section.

2. New algorithm

In this section, we present a new DQN method by modifying the subproblem (10). Andrei's motivation for introducing $\mathrm{tr}(B_k + \Delta_k)$ into the objective function of (10) was to obtain an updated matrix B_{k+1} with clustered eigenvalues, which improves its condition number. In this sense, in view of the fact that $\mathrm{tr}(B_k + \Delta_k) = \mathrm{tr}(B_k) + \mathrm{tr}(\Delta_k)$, it follows that $\mathrm{tr}(B_k)$ in the subproblem (10) does

not play any role throughout the minimization procedure, while the updated matrix B_{k+1} is a combination of B_k and Δ_k. To control the eigenvalues of matrix B_{k+1}, we modify the subproblem (10) as

$$\min_{\Delta_k \in \mathfrak{D}} \frac{1}{2}\|\Delta_k\|_F^2 + \mathrm{tr}(\Delta_k B_k) \quad \text{subject to} \quad s_k^\top (B_k + \Delta_k) s_k = s_k^\top y_k. \tag{12}$$

The constrained problem (12) can be solved by finding a stationary point of the Lagrangian function

$$L(\Delta_k, \lambda) = \frac{1}{2}\mathrm{tr}(\Delta_k^\top \Delta_k) + \mathrm{tr}(\Delta_k B_k) + \lambda(s_k^\top B_k s_k + s_k^\top \Delta_k s_k - s_k^\top y_k),$$

where λ is the Lagrange multiplier. In this sense, we obtain a solution of the nonlinear system

$$\frac{\partial L}{\partial \Delta_k} = \Delta_k + B_k + \lambda E_k^2 = 0, \quad \frac{\partial L}{\partial \lambda} = s_k^\top B_k s_k + s_k^\top \Delta_k s_k - s_k^\top y_k = 0, \tag{13}$$

where E_k is defined by (8), and subsequently Δ_k is obtained as

$$\Delta_k = -B_k - \lambda E_k^2. \tag{14}$$

Multiplying the relation (14) by $s_k^\top$ from the left and by s_k from the right and invoking equation (13), we can result in

$$-s_k^\top B_k s_k - \lambda s_k^\top E_k^2 s_k = s_k^\top y_k - s_k^\top B_k s_k,$$

and so

$$\lambda = -\frac{s_k^\top y_k}{s_k^\top E_k^2 s_k}. \tag{15}$$

Substituting (15) into (14) concludes that

$$\Delta_k = \frac{s_k^\top y_k}{s_k^\top E_k^2 s_k} E_k^2 - B_k,$$

and hence,

$$B_{k+1} = \theta_k E_k^2,$$

where

$$\theta_k = \frac{s_k^\top y_k}{s_k^\top E_k^2 s_k} = \frac{s_k^\top y_k}{\sum_{j=1}^n (s_k^{(j)})^4}.$$

Therefore, the search direction can be obtained as

$$d_{k+1} = -B_{k+1}^{-1} g_{k+1},$$

i.e., in components

$$d_{k+1}^{(i)} = -\frac{g_{k+1}^{(i)}}{b_{k+1}^{(i)}}, \quad i = 1, \ldots, n,$$

where

$$b_{k+1}^{(i)} = \theta_k (s_k^{(i)})^2, \quad i = 1, \ldots, n. \tag{16}$$

To be sure of the descent property of the search direction d_{k+1}, it is sufficient to have $b_{k+1}^{(i)} > 0$ for all $i = 1, \ldots, n$. In this sense, after computing $b_{k+1}^{(i)}$ from (16), we modify it as

$$\hat{b}_{k+1}^{(i)} = \begin{cases} b_{k+1}^{(i)} & \text{if } \varepsilon_b \leqslant b_{k+1}^{(i)} \leqslant \gamma, \\ 1 & \text{otherwise,} \end{cases} \tag{17}$$

where $\varepsilon_b \in (0, 1)$ and γ is a suitable upper bound for the diagonal elements. Theoretically, the roles of parameters γ and ε_b are to truncate the elements $b_{k+1}^{(i)}$ from two sides when they are too large or too small. As a result, the sequence $\{\widehat{B}_{k+1}\}$ always satisfies

$$\varepsilon_b \leqslant \hat{b}_{k+1}^{(i)} \leqslant \gamma, \quad i = 1, \ldots, n, \tag{18}$$

and is bounded. Based on this discussion, we define the search direction

$$d_{k+1} = -\widehat{B}_{k+1}^{-1} g_{k+1}, \tag{19}$$

i.e., in components

$$d_{k+1}^{(i)} = -\frac{g_{k+1}^{(i)}}{\hat{b}_{k+1}^{(i)}}, \quad i = 1, \ldots, n, \tag{20}$$

where $\hat{b}_{k+1}^{(i)}$ is defined by (17). From the above discussion, the next lemma follows.

Lemma 1. *Let $\hat{b}_{k+1}^{\min}$ and $\hat{b}_{k+1}^{\max}$ be the smallest and largest diagonal component of $\widehat{B}_{k+1}$, respectively. Then the direction d_{k+1} yielded by* (19) *satisfies*

$$|g_{k+1}^\top d_{k+1}| \geqslant \frac{\|g_{k+1}\|^2}{\hat{b}_{k+1}^{\max}}, \tag{21}$$

$$\frac{|g_{k+1}^\top d_{k+1}|}{\|d_{k+1}\|^2} \geqslant \frac{(\hat{b}_{k+1}^{\min})^2}{\hat{b}_{k+1}^{\max}}. \tag{22}$$

P r o o f. Clearly, we know that the inequality

$$\frac{1}{\hat{b}_{k+1}^{\max}} I_{n\times n} \leqslant \widehat{B}_{k+1}^{-1} \leqslant \frac{1}{\hat{b}_{k+1}^{\min}} I_{n\times n}$$

holds. Multiplying this inequality by $g_{k+1}^{\top}$ from the left and by g_{k+1} from the right results in

$$\frac{\|g_{k+1}\|^2}{\hat{b}_{k+1}^{\max}} \leqslant g_{k+1}^{\top} \widehat{B}_{k+1}^{-1} g_{k+1} \leqslant \frac{\|g_{k+1}\|^2}{\hat{b}_{k+1}^{\min}}. \tag{23}$$

Combining (23) with (19) implies the inequalities (21) and (22) immediately. □

Based on the above discussion, the new algorithm can be introduced.

Algorithm 1 (ADQN Algorithm).

Input: An initial point $x_0 \in \mathbb{R}^n$, positive constants ε_b, γ, and $\sigma, \beta \in (0,1)$.

Step 0. Set k = 0 and $d_0 = -g_0$.

Step 1. If the stop criterion is satisfied, stop. Otherwise, go to step 2.

Step 2. Set $x_{k+1} = x_k + t_k d_k$ with $t_k = \beta^m$, where m is the smallest non-negative integer such that t_k satisfies (3).

Step 3. Compute the search direction d_{k+1} by (20).

Step 4. Set $k = k + 1$ and go to step 1.

3. Convergence analysis

In this part, the global convergence of Algorithm 1 is proved. In this regard, we require the following assumptions.

Assumption 1.

(A1) The level set $L_0 = \{x \in \mathbb{R}^n;\ f(x) \leqslant f(x_0)\}$ is bounded.

(A2) In a neighborhood U of L_0, the function f is continuously differentiable and its gradient is Lipschitz continuous; namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leqslant L\|x - y\| \quad \forall\, x, y \in U.$$

The next lemma displays that at each iteration, the reduction of function f is at least as much as a constant multiple of $\|g_k\|^2$.

Lemma 2. *Suppose that $\{x_k\}$ is generated by Algorithm* 1 *and Assumption* 1 *is fulfilled. Then, there exists a constant $\mu > 0$ such that*

$$f_k - f_{k+1} \geqslant \mu \|g_k\|^2, \tag{24}$$

where $f_k = f(x_k)$.

P r o o f. According to step 2 of Algorithm 1, we know that $t_k \leqslant 1$ and

$$f_k - f_{k+1} \geqslant \sigma t_k |g_k^\top d_k| \quad \forall\, k \in \mathbb{N}. \tag{25}$$

Note that if $t_k = 1$, then from the relations (18) and (21), we conclude

$$f_k - f_{k+1} \geqslant \frac{\sigma}{\gamma}\|g_k\|^2. \tag{26}$$

Now, we consider $t_k < 1$. In this case, similarly to the proof of Proposition 1 in [1], we can deduce that

$$t_k > \frac{\beta(1-\sigma)}{L}\frac{|g_k^\top d_k|}{\|d_k\|^2}. \tag{27}$$

It follows from the relations (22) and (27) that

$$t_k > \frac{\beta(1-\sigma)}{L}\frac{(\hat{b}_k^{\min})^2}{\hat{b}_k^{\max}}. \tag{28}$$

Embedding the relations (18), (21), and (28) into (25), we get

$$f_k - f_{k+1} \geqslant \frac{\sigma\varepsilon_b^2(1-\sigma)}{L\gamma^2}\beta\|g_k\|^2. \tag{29}$$

Finally, by defining $\mu = \min\{\sigma/\gamma, \sigma\varepsilon_b^2(1-\sigma)\beta/(L\gamma^2)\}$, the inequalities (26) and (29) imply that

$$f_k - f_{k+1} \geqslant \mu\|g_k\|^2.$$

□

In the next theorem, the global convergence of Algorithm 1 with Armijo line search is proved.

Theorem 1. *Assume that $\{x_k\}$ is generated by Algorithm 1. Under Assumption 1, we have*

$$\lim_{k\to\infty} \|g_k\| = 0.$$

P r o o f. We can assume $f_{k+1} < f_k$ for any $k \geqslant 0$, since otherwise according to the relation (24) it follows $f_k = f_{k+1}$ and thus $g_k = 0$, which is the same as saying that x_k is a stationary point of the problem (1). On the other hand, according to Assumption 1, we know that the function f over the set L_0 is bounded from below, so the sequence $\{f_k\}$ is convergent and consequently

$$\lim_{k\to\infty} (f_k - f_{k+1}) = 0.$$

This fact along with the relation (24) imply that

$$\lim_{k\to\infty} \|g_k\| = 0,$$

and the proof is complete. □

4. Numerical experiments

This section reports some numerical behavior of Algorithm 1 to evaluate its numerical performance. We give some numerical results on a set of test problems consisting of 101 instances given by Andrei [2], in which most of the problems are extracted from the CUTE collection by Bongartz et al. [8]. We compare the ADQN method against the following DQN algorithms:

- ▷ SD: The steepest descent method. Consider Algorithm 1 with $d_{k+1} = -g_{k+1}$.
- ▷ DQN-B: A modified version of the DQN method proposed by Leong et al. [17]. Consider Algorithm 1 with $d_{k+1} = -B_{k+1}^{-1} g_{k+1}$ where B_{k+1} is given by (9).
- ▷ DNRTR: The DQN method proposed by Andrei [3]. Consider Algorithm 1 with $d_{k+1} = -B_{k+1}^{-1} g_{k+1}$ where B_{k+1} is given by (11).

All the codes were written in MATLAB R2018b and run on a Laptop with a 1.8 GHz CPU, 4 GB RAM and Windows 10 operation system.

For four tested algorithms, the step length t_k was chosen by the backtracking procedure with the parameters $\sigma = 0.1$ and $\beta = 0.5$. The following same parameters are used for all the algorithms:

$$\varepsilon_b = 0.01, \quad \gamma = 10^5.$$

We terminate algorithms when $\|g_k\| \leqslant 10^{-4}(1 + |f_k|)$ or the number of iterations exceeds 10000.

The efficiency of algorithms is investigated by using the performance profile introduced by Dolan and Moré [13]. They introduced the notion of a performance profile as a means to compare and evaluate the efficiency of the set of solvers S on a test set P. In this process, the profile of each algorithm is measured considering the ratio of its computational outcome versus the best numerical outcome of all algorithms. For n_s solvers and n_p problems, the performance profile $\varrho\colon \mathbb{R} \to [0, 1]$ is defined as follows:

For each problem $p \in P$ and solver $s \in S$, $t_{p,s}$ is the computation result regarding the performance index. Then, the performance ratio is defined by

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s}\,;\ s \in S\}},$$

and the performance profile is given by

$$\varrho(\tau) = \frac{1}{n_p} \text{ size } \{p \in P\,;\ r_{p,s} \leqslant \tau\} \quad \forall\, \tau \in \mathbb{R},$$

where size$\{C\}$ is the number of elements in the set C. Note that $\varrho(\tau)$ is the probability for solver $s \in S$ that a performance ratio $r_{p,s}$ is within a factor $\tau \in \mathbb{R}$ of the best possible ratio. The function ϱ is the cumulative distribution one for the performance ratio. In the figures, the top curve is the method that solved the most problems in a time that was within a factor τ of the best time [13]. The left side of the figures shows the percentage of the test problems for which a method is the fastest. The right side of the profile shows the percentage of the test problems that are successfully solved by each of the algorithms [13].

We examine algorithms based on the number of function evaluations (N_f) and gradient evaluations (N_g). Figures 1 and 2 indicate the performance of the above mentioned algorithms based on N_f and N_g for $n = 1008$, respectively.

With a glimpse into these figures, the following results are deduced.

Figure 1 displays that the ADQN method solves nearly 43% of the test problems at least N_f evaluations, while DNRTR, DQN-B and SD methods solve nearly 36%, 37%, and 22% of the test problems, respectively. Figure 2 displays that the ADQN method solves nearly 46% of the test problems at least N_g evaluations, while DNRTR, DQN-B and SD methods solve nearly 31%, 33%, and 26% of the test problems, respectively.

According to these results, we can result that the ADQN method has performed better than other methods.

5. Conclusions

We have presented a new DQN method for solving unconstrained optimization problems. The proposed method utilizes new criteria to generate the Hessian approximation, which effectively controls the diagonal elements. Additionally, the global convergence of the proposed method was shown by the Armijo line search. Numerical results on a collection of standard test problems indicated that the proposed method outperforms over several existing DQN methods. These findings suggest that the ADQN method has potential for effectively solving large-scale optimization problems.

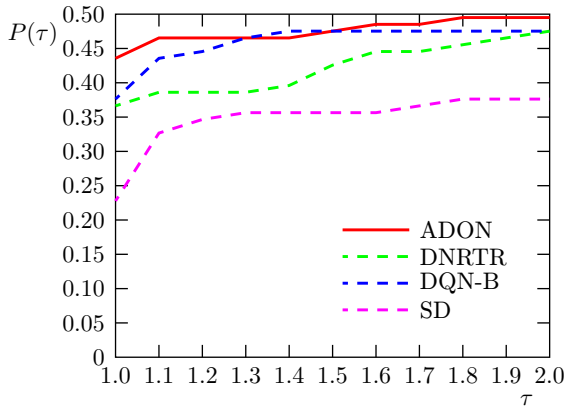

Figure 1. Performance profile of the algorithms based on N_f.

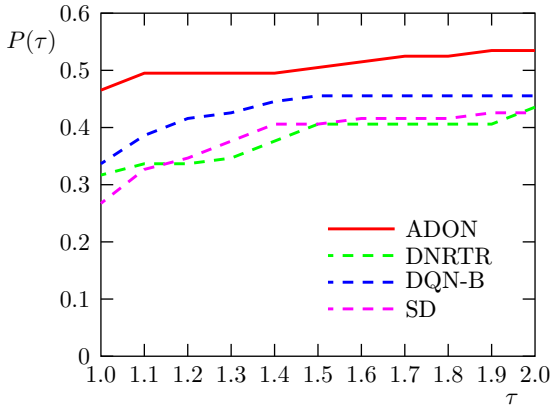

Figure 2. Performance profile of the algorithms based on N_g.

References

[1] *N. Andrei*: An acceleration of gradient descent algorithm with backtracking for unconstrained optimization. Numer. Algorithms *42* (2006), 63–73. zbl MR doi

[2] *N. Andrei*: An unconstrained optimization test functions collection. Adv. Model. Optim. *10* (2008), 147–161. zbl MR

[3] *N. Andrei*: A diagonal quasi-Newton updating method based on minimizing the measure function of Byrd and Nocedal for unconstrained optimization. Optimization *67* (2018), 1553–1568. zbl MR doi

[4] *N. Andrei*: A diagonal quasi-Newton updating method for unconstrained optimization. Numer. Algorithms *81* (2019), 575–590. zbl MR doi

[5] *N. Andrei*: A new accelerated diagonal quasi-Newton updating method with scaled forward finite differences directional derivative for unconstrained optimization. Optimization *70* (2021), 345–360. zbl MR doi

[6] *L. Armijo*: Minimization of functions having Lipschitz continuous first partial derivatives. Pac. J. Math. *16* (1966), 1–3. zbl MR doi

[7] *J. Barzilai, J. M. Borwein*: Two-point step size gradient methods. IMA J. Numer. Anal. *8* (1988), 141–148. zbl MR doi
[8] *I. Bongartz, A. R. Conn, N. Gould, P. L. Toint*: CUTE: Constrained and unconstrained testing environment. ACM Trans. Math. Softw. *21* (1995), 123–160. zbl doi
[9] *C. G. Broyden, J. E. Dennis,* Jr., *J. J. Moré*: On the local and superlinear convergence of quasi-Newton methods. J. Inst. Math. Appl. *12* (1973), 223–245. zbl MR doi
[10] *J. E. Dennis,* Jr., *J. J. Moré*: A characterization of superlinear convergence and its application to quasi-Newton methods. Math. Comput. *28* (1974), 549–560. zbl MR doi
[11] *J. E. Dennis,* Jr., *J. J. Moré*: Quasi-Newton methods, motivation and theory. SIAM Rev. *19* (1977), 46–89. zbl MR doi
[12] *J. E. Dennis,* Jr., *H. Wolkowicz*: Sizing and least-change secant methods. SIAM J. Numer. Anal. *30* (1993), 1291–1314. zbl MR doi
[13] *E. D. Dolan, J. J. Moré*: Benchmarking optimization software with performance profiles. Math. Program. *91* (2002), 201–213. zbl MR doi
[14] *M. Farid, W. J. Leong, L. Zheng*: A new diagonal gradient-type method for large scale unconstrained optimization. Sci. Bull., Ser. A, Appl. Math. Phys., Politeh. Univ. Buchar. *75* (2013), 57–64. zbl MR
[15] *P. E. Gill, W. Murray*: Conjugate-gradient methods for large-scale nonlinear optimization. Technical Report SOL-79-15. Stanford University, Stanford, 1979, pp. 1–66.
[16] *A. A. Goldstein*: On steepest descent. J. Soc. Ind. Appl. Math., Ser. A: Control *3* (1965), 147–151. zbl MR doi
[17] *W. J. Leong, S. Enshaei, S. L. Kek*: Diagonal quasi-Newton methods via least change updating principle with weighted Frobenius norm. Numer. Algorithms *86* (2021), 1225–1241. zbl MR doi
[18] *W. J. Leong, M. Farid, M. A. Hassan*: Improved Hessian approximation with modified quasi-Cauchy relation for a gradient-type method. Adv. Model. Optim. *12* (2010), 37–44. zbl MR
[19] *S. G. Nash*: Preconditioning of truncated-Newton methods. SIAM J. Sci. Stat. Comput. *6* (1985), 599–616. zbl MR doi
[20] *J. Nocedal*: Updating quasi-Newton matrices with limited storage. Math. Comput. *35* (1980), 773–782. zbl MR doi
[21] *J. Nocedal, S. J. Wright*: Numerical Optimization. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006. zbl MR doi
[22] *M. J. D. Powell*: A new algorithm for unconstrained optimization. Nonlinear Programming. Elsevier, Amsterdam, 1970, pp. 31–65. zbl MR doi
[23] *M. Raydan*: The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM J. Optim. *7* (1997), 26–33. zbl MR doi
[24] *P. Wolfe*: Convergence conditions for ascent methods. SIAM Rev. *11* (1969), 226–235. zbl MR doi
[25] *M. Zhu, J. L. Nazareth, H. Wolkowicz*: The quasi-Cauchy relation and diagonal updating. SIAM J. Optim. *9* (1999), 1192–1204. zbl MR doi

Authors' address: *Mahsa Nosrati*, *Keyvan Amini* (corresponding author), Department of Mathematics, Faculty of Science, Razi University, Taq-e Bostan, 6714414971 Kermanshah, Iran, e-mail: `m.nosrati@razi.ac.ir`, `kamini@razi.ac.ir`.