Computers and Mathematics with Applications 59 (2010) 3301–3307

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

A new two-step gradient-type method for large-scale unconstrained optimization

Mahboubeh Farid *, Wah June Leong, Malik Abu Hassan

Institute for Mathematical Research, University Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 8 August 2009
Received in revised form 23 February 2010
Accepted 11 March 2010

Keywords:
Diagonal updating
Weak secant equation
Two-step gradient method
Barzilai and Borwein method

ABSTRACT

In this paper, we propose some improvements on a new gradient-type method for solving large-scale unconstrained optimization problems, in which we use data from two previous steps to revise the current approximate Hessian. The new method which we considered, resembles to that of Barzilai and Borwein (BB) method. The innovation features of this approach consist in using approximation of the Hessian in diagonal matrix form based on the modified weak secant equation rather than the multiple of the identity matrix in the BB method. Using this approach, we can obtain a higher order accuracy of Hessian approximation when compares to other existing BB-type method. By incorporating a simple monotone strategy, the global convergence of the new method is achieved. Practical insights into the effectiveness of the proposed method are given by numerical comparison with the BB method and its variant.

1. Introduction

In this paper, we focus on the minimization of a smooth function of n variables:

$$\min f(x), \tag{1}$$

where $x \in R^n$ and n is large. The Barzilai and Borwein (BB) gradient method for solving (1) takes the form

$$x_{k+1} = x_k - B_k^{-1} g_k, \tag{2}$$

where $g_k = \nabla f(x_k)$ and the stepsize α_k is determined by letting $B_k = (1/\alpha_k)I$ to be an approximation of the Hessian of f at x_k. Let us denote $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$. In order for B_{k+1}, the update of B_k, to have a certain quasi-Newton property

$$B_{k+1} s_k = y_k. \tag{3}$$

Barzilai and Borwein [1] choose α_k such that

$$B_{k+1} = \arg\min_{B=(1/\alpha)I} \|B_{k+1} s_k - y_k\|_2, \tag{4}$$

and yields

$$\alpha_{k+1} = \frac{s_k^T y_k}{s_k^T s_k}. \tag{5}$$

* Corresponding author. Tel.: +60 176032604; fax: +60 389423789.
E-mail addresses: mahboubeh@inspem.upm.edu.my, mahboubeh_farid@yahoo.com (M. Farid), leong@math.upm.edu.my (W.J. Leong), malik@fsas.upm.edu.my (M.A. Hassan).

0898-1221/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2010.03.014

The BB gradient method is preferable over the classic steepest descent method both in theory and in real computations. It is known that the classical steepest descent method converges linearly and is badly effected by ill conditioning (see [2]). On the other hand, the BB gradient method does not require line search and often needs less computation. Moreover, BB method is shown to be *R*-superlinearly convergent for two-dimensional strictly convex quadratics [1] and for any dimensional convex quadratics, the method is shown to be *R*-linearly convergent [3]. Due to its simplicity and numerical efficiency, the BB gradient method has now received many attentions, for example see [4–6]. BB method greatly speeds up the convergence, however it does not guarantee descent on the objective function at every iteration and it is not easy to generalize the method to general nonlinear functions.

To cope with these advantages, a fixed-step gradient-type method of BB-kind were suggested by Hassan et al. [7] and Leong et al. [8], respectively. Leong et al. [8] considered the approximation of the Hessian by a diagonal matrix based on weak secant (or weak quasi-Newton) equation

$$s_k^T B_{k+1} s_k = s_k^T y_k. \tag{6}$$

Their gradient algorithm is monotone in the sense that it guarantees descent in each iteration while function and extra gradient evaluations are not required. Global and a linear rate of convergence are also established for their method.

On the other hand, Hassan et al. [7] derived a diagonal updating formula for approximation the inverse Hessian under similar approach of Leong et al. [8]. Incorporating a different strategy, they showed that their algorithm can ensure descent on function values in each iteration and is also globally converged. Apart from these advances, both of these diagonal updating employ a standard one-step two-point approach in Hessian approximation, which is commonly used by standard BB or quasi-Newton methods. In contrast, this paper develops a new gradient-type method for solving large-scale unconstrained optimization problem by extending and adapting the approach of [7,8] while uses a two-step multi-point approach to increase the accuracy of Hessian approximation. The fundamental idea in here is to furnish an interpolating curve in a variable space and later uses it to derive a generalized weak secant equation, which can then be employed in the construction of our new Hessian approximation.

The rest of this paper is organized as follows: In Section 2, we present a new diagonal updating scheme for gradient-type method based on the improved weak secant equation. Section 3 discusses the properties of this new algorithm and a new monotone gradient algorithm is described. The global convergent under mild assumption will be established in Section 4 and numerical results are reported in Section 5.

2. Two-step diagonal updates

In this section we define our new gradient method based on diagonal updating. This method generates a sequence of points $\{x_k\}$ by

$$x_{k+1} = x_k - B_k^{-1} g_k, \tag{7}$$

where B_k is a diagonal matrix. Our aim is to construct a matrix B_k through diagonal updating which is a good approximation of the Hessian. Note that any approximation of Hessian should satisfy the secant equation (3) to ensure that the curvature information is corrected. However, since B_k is diagonal and to maintain only $O(n)$ floating point operation, it is reasonable to let B_k satisfies only the weak secant equation (6). By using only s_k and y_k, we just utilize data from one previous step to revise the current approximate Hessian and it is known as one-step gradient method. In this section, in order for B_k to be a more accurate approximation of the Hessian matrix, we employ an interpolating curve in the variable-space to derive an approximation generalization of weak secant equation. We consider one of the most successful of two-step method which employs the current approximation to the Hessian to determine the parameterizations of the interpolating curve and, hence, the derivatives which are required in the generalized updating formula. (see for more detail in [9–12]).
Now, by incorporating this two-step information we can offer an improved weak secant equation as follows:

$$(s_k - \gamma_k s_{k-1})^T B_{k+1}(s_k - \gamma_k s_{k-1}) = (s_k - \gamma_k s_{k-1})^T (y_k - \gamma_k y_{k-1}), \tag{8}$$

or say

$$r_k^T B_{k+1} r_k = r_k^T w_k. \tag{9}$$

In formula (8) and (9) we use data from the last two steps instead of one previous step. Therefore it is needed to construct interpolating quadratic curves $x(\tau)$ and $h(\tau)$ (where $\tau \in R$) such that $x(\tau)$ interpolates the two last iterates x_{k-1}, x_k and x_{k+1}, and $h(\tau)$ interpolates the corresponding last gradient vectors g_{k-1}, g_k and g_{k+1} (which are assumed to be available). Since we are using more information from several recent steps in the construction of $x(\tau)$ (by contrast, the derivation of standard BB and quasi-Newton methods utilizes only data from the most recent step, i.e. x_k to x_{k+1}), we expect that the new secant equation derived from this approach, will also be able to improve the accuracy in Hessian approximation. The derivatives of these two curves at $\tau = \tau_2$, where τ_j is the value of τ for which

$$x(\tau_j) = x_{k+j-1}, \quad j = 0, 1, 2 \tag{10}$$

$$h(\tau_j) = g_{k+j-1}, \quad j = 0, 1, 2 \tag{11}$$

are then substituted into the relation (derived from applying the Chain Rule to $g(x(\tau))$),

$$G(x_{k+1})x'(\tau_2) = g'(x(\tau_2)), \tag{12}$$

where primes denote differentiation with respect to τ. Note that

$$r_k \overset{\text{def}}{=} x'(\tau_2), \tag{13}$$

and

$$w_k \overset{\text{def}}{=} h'(\tau_2) \approx g'(x(\tau_2)). \tag{14}$$

To determine γ_k in (8), we need the values of τ_0, τ_1 and τ_2. Here we consider one of the successful approach that was introduced by Ford and Moghrabi [11]: consider a norm of the general form $\|z\|_M \overset{\text{def}}{=} \{z^T M z\}^{1/2}$ where M is symmetric-positive-definite matrix and here we let $M = B_k$. Without the loss of generality, we choose $\tau_2 = 0$ and the value of $\{\tau_j\}_{j=0}^2$ is determined as follows:

$$\begin{aligned} -\tau_1 &= \tau_2 - \tau_1 \\ &\overset{\text{def}}{=} \|x(\tau_2) - x(\tau_1)\|_{B_k} \\ &= \|x_{k+1} - x_k\|_{B_k} \\ &= \|s_k\|_{B_k} \\ &= \left(s_k^T B_k s_k\right)^{1/2}, \end{aligned} \tag{15}$$

and

$$\begin{aligned} -\tau_0 &= \tau_2 - \tau_0 \\ &\overset{\text{def}}{=} \|x(\tau_2) - x(\tau_0)\|_{B_k} \\ &= \|x_{k+1} - x_{k-1}\|_{B_k} \\ &= \|s_k + s_{k-1}\|_{B_k} \\ &= \left((s_k + s_{k-1})^T B_k (s_k + s_{k-1})\right)^{1/2}. \end{aligned} \tag{16}$$

By defining the quantity δ by the ratio

$$\delta = \frac{\tau_2 - \tau_1}{\tau_1 - \tau_0}, \tag{17}$$

r_k, w_k are given by the following expressions

$$r_k = s_k - \frac{\delta^2}{1 + 2\delta} s_{k-1}, \tag{18}$$

$$w_k = y_k - \frac{\delta^2}{1 + 2\delta} y_{k-1}. \tag{19}$$

Hence, when the inequality $r_k^T w_k > 10^{-4} \|r_k\|_2 \|w_k\|_2$ is satisfied, we can state the resulting update formula for B_{k+1} in the following theorem,

Theorem 2.1. *Assume that $B_k > 0$ is a positive definite diagonal matrix and B_{k+1} is an updated version of B_k, which is also diagonal. Let us denote $\Delta_k = B_{k+1} - B_k$ as the deviation between B_{k+1} and B_k. Suppose that $r_k \neq 0$ where r_k and w_k are defined by (18) and (19), respectively. Consider the following minimization problem:*

$$\begin{aligned} &\min \frac{1}{2} \|\Delta_k\|_F^2, \\ &\text{s.t. } r_k^T \Delta_k r_k = r_k^T w_k - r_k^T B_k r_k, \\ &\text{and } \Delta_k \text{ is a diagonal matrix,} \end{aligned} \tag{20}$$

where $\|.\|_F$ denotes the Frobenius norm. Then the optimal solution of (20) is given by

$$\Delta_k^{(i)} = \frac{\left(r_k^T w_k - r_k^T B_k r_k\right)}{\sum_{i=1}^{n} (r_k^{(i)})^4} (r_k^{(i)})^2, \quad i = 1, 2, \ldots, n \tag{21}$$

where the $\Delta_k^{(i)}$ is ith diagonal component of Δ_k and $r_k^{(i)}$ is the ith component of vector r_k.

Proof. In (20) the objective function is strictly convex and the feasible set is also convex. Therefore we can state the Lagrangian function for (20) as follows

$$L(\Delta_k, \mu) = \frac{1}{2}\|\Delta_k\|_F^2 + \mu\left(r_k^T\Delta_k r_k - r_k^T w_k + r_k^T B_k r_k\right) \tag{22}$$

where μ is the Lagrange multiplier associated with constraint. Differentiating L with respect to each component of Δ_k and setting them to zero yields

$$\Delta_k^{(i)} = -\mu(r_k^{(i)})^2 \quad \text{for all } i = 1, 2, \ldots, n. \tag{23}$$

By multiplying each of Eq. (23) by $(r_k^{(i)})$ for each $i = 1, 2, \ldots, n$, adding them together and invoking the constraint $r_k^T\Delta_k r_k = r_k^T w_k - r_k^T B_k r_k$ we have

$$\mu = \frac{(r_k^T w_k - r_k^T B_k r_k)}{\sum_{i=1}^{n}(r_k^{(i)})^4}. \tag{24}$$

Finally, by substituting (24) into (23) we obtain (21). □

It follows from Theorem 2.1 that the optimal updating formula for B_{k+1} is given by

$$b_{k+1}^{(i)} = b_k^{(i)} + \frac{(r_k^T w_k - r_k^T B_k r_k)}{\sum_{i=1}^{n}(r_k^{(i)})^4}(r_k^{(i)})^2, \quad i = 1, 2, \ldots, n \tag{25}$$

where $b_{k+1}^{(i)}$ and $b_k^{(i)}$ is the ith diagonal element of B_{k+1} and B_k, respectively. However there is no guarantee that either $B_{k+1} > 0$ or the sequence $\{f(x_k)\}$ generated by gradient-type method with updating scheme (7) is always monotonically decreased. Hence some safeguards are introduced.

In the next section, we will include safeguards to maintain positive-definiteness of B_{k+1} that based on scaling strategy and also strategy for preserving monotonicity in function values.

3. Scaling on the diagonal updates

In order to define our safeguarding strategy, we first describe the scaled diagonal update. The scaled diagonal updating formula is similar to formula (25), except that the matrix B_k is replaced by $\sigma_k B_k$:

$$\overline{B}_{k+1} = B_{k+1}(\sigma_k B_k, r_k, w_k), \tag{26}$$

where σ_k is a scaling parameter. Note that $\overline{B}_{k+1}$ still satisfies the weak secant equation.
From (25), we observe that the possibility of non positive-definiteness occurs in B_{k+1} is when $r_k^T w_k - r_k^T B_k r_k < 0$. Therefore by considering scaling parameter

$$\sigma_k = \min\left(\frac{r_k^T w_k}{r_k^T B_k r_k}, 1\right), \tag{27}$$

we have

$$\begin{aligned} b_{k+1}^{(i)} &= \sigma_k b_k^{(i)} + \frac{(r_k^T w_k - r_k^T \sigma_k B_k r_k)}{\sum_{i=1}^{n}(r_k^{(i)})^4}(r_k^{(i)})^2 \\ &= \begin{cases} \dfrac{r_k^T w_k}{r_k^T B_k r_k} b_k^{(i)} & \text{if } r_k^T w_k - r_k^T B_k r_k < 0 \\ b_k^{(i)} + \dfrac{(r_k^T w_k - r_k^T B_k r_k)}{\sum_{i=1}^{n}(r_k^{(i)})^4}(r_k^{(i)})^2 & \text{if } r_k^T w_k - r_k^T B_k r_k > 0 \end{cases} \quad i = 1, 2, \ldots, n \end{aligned} \tag{28}$$

where r_k, w_k and σ_k is given by (18), (19) and (27) respectively. Hence we can guarantee the positive-definiteness of B_{k+1}. Moreover scheme (28) cannot guarantee monotonicity of function value. One can show that if

$$b_{k+1}^M < \frac{2(b_k^m)^2}{b_k^M} \tag{29}$$

where b_k^m, b_k^M, b_{k+1}^M be the smallest and largest diagonal component of B_k and B_{k+1}, respectively, then the possibility of non-monotonic behavior occurs in the sequence $\{f(x_k)\}$ might be observed. Hence a safeguard on this limitation is needed.

In the following, we present our algorithm as well as the safeguard strategy for monotonicity in the detail:

TMDGrad Algorithm.

Step 0. Choose an initial point $x_0 \in R^n$, and a positive definite matrix $B_0 = I$. Let $k := 0$.
Step 1. Compute g_k. If $\|g_k\| \leq \epsilon$, stop.
Step 2. If $k = 0$, set $x_1 = x_0 - \frac{g_0}{\|g_0\|}$, Else if $k = 1$ set $r_k = s_k$ and $w_k = y_k$ go to Step 4.
Step 3. For $k \geq 2$, compute τ_0, τ_1 and δ, by Eqs. (15)–(17), respectively, calculate r_k, w_k and σ_k by Eqs. (18), (19) and (27), respectively.
If $r_k^T w_k \leq 10^{-4} \|r_k\|_2 \|w_k\|_2$,
set $r_k = s_k$ and $w_k = y_k$.
Step 4. Let $x_{k+1} = x_k - B_k^{-1} g_k$ and update $B_{k+1} = \operatorname{diag}\left(b_k^{(1)}, b_k^{(2)}, \ldots, b_k^{(n)}\right)$, where $b_k^{(i)}$, $i = 1, 2, \ldots, n$ is given by (28).
Step 5. Check whether $b_{k+1}^M < \frac{2(b_k^m)^2}{b_k^M}$:
If yes, set $B_{k+1} = \vartheta I$ where $\vartheta = \min\left(\frac{b_k^M}{2(b_k^m)^2}, \frac{\delta_k^T \gamma_k}{\delta_k^T \delta_k}\right)$.
Else retain B_{k+1} that is computed by Step 4.
Step 6. Set $k := k + 1$ and return to Step 1.

4. Convergence analysis

We shall also establish the convergence of the TMDGrad algorithm when applied to the minimization of a strictly convex quadratic function with constant Hessian.

Theorem 4.1. *Assume that $f(x) \in C^2$ is a strictly convex quadratic function. Let $\{x_k\}$ be a sequence generated by the TMDGrad method and x^* is a unique minimizer of f. Then either $g_k = 0$ holds for some finite $k \geq 1$, or $\lim_{k\to\infty} \|g_k\| = 0$.*

Proof. Denote $G = \nabla^2 f$. Again let b_k^m, b_k^M, b_{k+1}^m and b_{k+1}^M be the smallest and largest diagonal elements of B_k and B_{k+1}, respectively. Consider the Taylor expansion of the strictly convex function, f at x_{k+1}:

$$f(x_k - B_k^{-1} g_k) = f(x_k) - g_k^T B_k^{-1} g_k + \frac{1}{2} g_k^T B_k^{-1} G B_k^{-1} g_k. \tag{30}$$

Since $Gr_k = w_k$, it follows that $r_k^T G r_k = g_k^T B_k^{-1} B_{k+1} B_k^{-1} g_k$.Thus

$$f(x_{k+1}) \leq f(x_k) - c \|g_k\|^2, \tag{31}$$

where $c = (b_k^M)^{-1} - \frac{(b_k^m)^{-2} b_{k+1}^M}{2} > 0$. If $c > 0$ $\left(\text{or } b_{k+1}^M > \frac{2(b_k^m)^2}{b_k^M}\right)$, we have $f(x_{k+1}) \leq f(x_k)$ for all k.

Else if $c < 0$ $\left(\text{or } b_{k+1}^M < \frac{2(b_k^m)^2}{b_k^M}\right)$, then we let $B_{k+1} = \vartheta I$ where $\vartheta = \min\left(\frac{b_k^M}{2(b_k^m)^2}, \frac{s_k^T y_k}{s_k^T s_k}\right)$. Hence (31) becomes

$$f(x_{k+1}) \leq f(x_k) - \bar{c} \|g_k\|^2,$$

where $\bar{c} = b_k^m - \left((b_k^M)^2 \vartheta\right)/2$. With our choice of ϑ, we have that $\bar{c} \geq 0$. This implies that $f(x_{k+1}) \leq f(x_k)$ for all k and since f is bounded below, it follows that

$$\lim_{k\to\infty} f(x_k) - f(x_{k+1}) = 0.$$

As $f(x_k) - f(x_{k+1}) \to 0$, and $c > 0$ then $\lim_{k\to\infty} \|g_k\| = 0$, i.e. x_k convergence to x^*. □

5. Numerical results

The TMDGrad method described in Section 3 is compared with the BB method and with MDGrad method. MDGrad method is implemented using MonoGrad of Leong et al. [8] with a different monotone strategy. All of these methods belong to a class of gradient methods without line searches.

The numerical experiments are carried out on a set of 30 test function given in Table 1 with dimension ranging from 10 to 10^4. The full description of these test problems can be found in [13,14].

All of the experiments are run on a PC with Core Duo CPU and the codes are written with Matlab 7.0. The stop criterion are $\|g_k\| \leq 10^{-4}$. We also force the routine to stop if the number of iteration exceed 1000. The performance of TMDGrad, MDGrad and BB method, relative to iteration, is evaluated using the profiles of Dolan and Moré (see [15]).

Table 1
Test problem and its dimension.

Problem	Dimension	References
Extended trigonometric, Penalty 1, Penalty 2, Quadratic QF2, Diagonal 4, Diagonal 5, Generalized tridiagonal 1, Generalized Rosenbrock, Generalized PSC1, Extended Himmelblau, Extended three exponential terms, Extended block diagonal BD1,	10, …,10 000	Moré et al. [14]
Extended PSC1, Raydan 2, Extended tridiagonal 2,	10, …,10 000	Andrei [13]
Extended Beale, Broyden tridiagonal, Quadratic diagonal perturbed, Perturbed quadratic, Quadratic QF1, Diagonal 1, Diagonal 2, Hager, Diagonal 3, Generalized Tridiagonal 2, Almost perturbed quadratic, Tridiagonal perturbed quadratic, Full Hessian FH1, Full Hessian FH2,	10, …,1000	Moré et al. [14]
Raydan 1,	10, …,1000	Andrei [13]

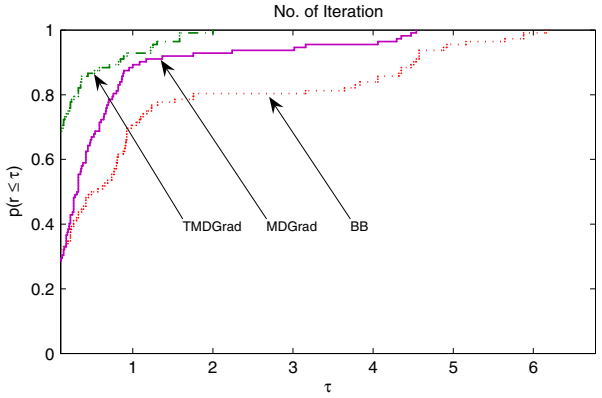

Fig. 1. Performance profile based on iterations.

From Fig. 1, we can see that, TMDGrad algorithm outperforms the MDGrad algorithm and BB method. In addition, our new method yields the best performance for large-scale unconstrained optimization. The numerical evidence provided by the tests are reported in Table 1 and illustrated in Fig. 1 demonstrates clearly that TMDGrad shows significant improvement, when compared with the BB and single-step MDGrad. Note that our algorithm still needs just $O(n)$ storage.

Especially it is different from other monotone gradient methods (for example see [16,17]) in the sense that function evaluation and line searches are not required. Numerical results on a large number of problems indicate that our methods are very promising.

6. Conclusion

A technique for update Hessian approximation in diagonal matrix form in two-step weak secant equation has been introduced. Numerical results illustrated in Fig. 1 demonstrate clearly that the TMDGrad method shows significant improvements, when compared with the BB and MDGrad method. Hence we can conclude that TMDGrad method is preferable due to encouraging numerical results, simple to implement, higher order accuracy of Hessian approximation, $O(n)$ storage requirement and globally converged.

Acknowledgements

The authors are very grateful to the anonymous referees for the useful suggestions.

References

[1] J. Barzilai, J.M. Borwein, Two point step size gradient methods, IMA J. Numer. Anal. 8 (1988) 141–148.
[2] H. Akaike, On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method, Ann. Inst. Statist. Math. 11 (1959) 1–17.
[3] Y.H. Dai, L.Z. Liao, *R*-linear convergence of the Barzilai and Borwein gradient method, IMA J. Numer. Anal. 22 (2002) 1–10.
[4] Z. Bin, L. Gao, Y.H. Dai, Monotone projected gradient methods for large-scale box-constrained quadratic programming, Sci. China Ser. A 49 (2006) 688–702.
[5] Y.H. Dai, R. Fletcher, Projected Barzilai–Borwein methods for large-scale box-constrained quadratic programming, Numer. Math. 100 (2005) 21–47.
[6] M. Raydan, On the Barzilai and Borwein choice of steplength for the gradient method, IMA J. Numer. Anal. 13 (1993) 618–622.
[7] M.A. Hassan, W.J. Leong, M. Farid, A new gradient method via quasi-Cauchy relation which guarantees descent, J. Comput. Appl. Math. 230 (2009) 300–305.

[8] W.J. Leong, M.A. Hassan, M. Farid, A monotone gradient method via weak secant equation for unconstrained optimization, Taiwanese J. Math. 14 (2) (2010) 413–423.
[9] J.A. Ford, Implicit updates in multi-step quasi-Newton methods, Comput. Math. Appl. 42 (2001) 1083–1091.
[10] J.A. Ford, L.A. Moghrabi, Multi-step quasi-Newton methods for optimization, J. Comput. Appl. Math. 50 (1994) 305–323.
[11] J.A. Ford, L.A. Moghrabi, Alternating multi-step quasi-Newton methods for unconstrained optimization, J. Comput. Appl. Math. 82 (1997) 105–116.
[12] J.A. Ford, S. Thrmlikit, New implicite updates in multi-step quasi-Newton methods for unconstrained optimization, J. Comput. Appl. Math. 152 (2003) 133–146.
[13] N. Andrei, An unconstrained optimization test functions collection, Adv. Model. Optim. 10 (2008) 147–161.
[14] J.J. Moré, B.S. Garbow, K.E. Hillstorm, Testing unconstrained optimization software, ACM Trans. Math. Softw. 7 (1981) 17–41.
[15] E.D. Dolan, J.J. Moré, Benchmarking optimization software with perpormance profiles, Math. Program. 91 (2002) 201–213.
[16] Y.H. Dai, J.Y. Yuan, Y. Yuan, Modified two-point stepsize gradient methods for unconstrained optimization, Comput. Optim. Appl. 22 (2002) 103–109.
[17] Y.H. Dai, Y. Yuan, Alternative minimization gradient method, IMA J. Numer. Anal. 23 (2003) 373–393.