

## A MONOTONE GRADIENT METHOD VIA WEAK SECANT EQUATION FOR UNCONSTRAINED OPTIMIZATION

Wah June Leong, Malik Abu Hassan and Mahboubeh Farid

**Abstract.** In this paper we present a new algorithm of steepest descent type. A new technique for steplength computation and a monotone strategy are provided in the framework of the Barzilai and Borwein method. In contrast with Barzilai and Borwein approach's in which the steplength is computed by means of a simple approximation of the Hessian in the form of scalar multiple of identity and an interpretation of the secant equation, the new proposed algorithm considers another approximation of the Hessian based on the weak secant equation. By incorporating a simple monotone strategy, the resulting algorithm belongs to the class of monotone gradient methods with linearly convergence. Numerical results suggest that for non-quadratic minimization problem, the new method clearly outperforms the Barzilai-Borwein method.

### 1. INTRODUCTION

Consider the problem:

$$(1.1) \quad \min f(x), \quad x \in R^n$$

where  $f$  is continuous and twice differentiable.

In this paper we are interested in the gradient method with updating scheme

$$(1.2) \quad x_{k+1} = x_k - \alpha_k g_k$$

where  $g_k = \nabla f(x_k)$  is the gradient of  $f(x)$  at the current iterate  $x_k$  and  $\alpha_k > 0$  is the steplength. The steepest descent (SD) method calculates the steplength by an exact line search

$$(1.3) \quad \alpha_k^{SD} = \arg \min_{\alpha \in R} f(x_k - \alpha g_k).$$

---

Received April 14, 2008, accepted May 29, 2008.

Communicated by B. S. Mordukhorich.

2000 *Mathematics Subject Classification*: 90C30, 65K05.

*Key words and phrases*: Unconstrained optimization, Monotone gradient methods, Weak secant equation, Barzilai-Borwein method.

It is well known that the SD method may converge very slowly, and thus being of no practical value.

In 1988 Barzilai and Borwein [1] proposed a gradient method (BB method) that uses a different strategy for choosing the steplength  $\alpha_k$ . The steplength along the negative gradient direction is computed from a two point approximation to the secant equation from quasi-Newton methods. They regard the matrix  $D_k = \gamma_k I$  as an approximation to the Hessian of  $f$  at  $x_k$ . In order for  $D_k$  to have correct curvature information, they impose some quasi-Newton properties on  $D_k$  and choose  $\gamma_k$  such that

$$(1.4) \quad \arg \min_{\gamma \in R} \|\gamma_k s_{k-1} - y_{k-1}\|_2$$

where  $s_{k-1} = x_k - x_{k-1}$  and  $y_{k-1} = g_k - g_{k-1}$ . Since in quasi-Newton method we have that

$$(1.5) \quad x_{k+1} = x_k - B_k^{-1} g_k,$$

and the quasi-Newton matrix  $B_k$  satisfies the secant equation

$$(1.6) \quad B_k s_{k-1} = y_{k-1}.$$

With these, Barzilai and Borwein [1] suggested

$$(1.7) \quad \gamma_k = \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}$$

and the method of BB is given by the following iterative scheme:

$$(1.8) \quad x_{k+1} = x_k - D_k^{-1} g_k,$$

where  $D_k^{-1}$  is given by  $\alpha_k^{BB} I$  and  $\alpha_k^{BB} = 1/\gamma_k$ .

For 2-dimensional convex quadratic function, Barzilai and Borwein [1] proves that the method (1.8) is  $R$ -superlinearly convergent. Having in view its simplicity, low storage requirement ( $3n$  locations for an  $n$ -dimensional problems) and numerical efficiency for well-conditioned problems, proved by Raydan [8] and Fletcher [6], the BB gradient has received a great deal of attention. For example, see [2], [3], [6], [9]. Despite all these advances, there are a number of reasons that show the inferior of the BB method (see Fletcher [6]). At first glance is the possibility of non-monotonic behavior in the sequence  $\{f(x_k)\}$  of the BB method, and the extent of the non-monotonicity depends in some way on the condition of the objective function. Also there is limited scope for the BB method to improve as regards with elapsed time. Therefore, it is the purpose of this paper to present a new algorithm of gradient type to address these drawbacks. Along this line, a new

gradient method is proposed in the next section. A simple monotone strategy is derived and the corresponding monotone gradient algorithm is described in Section 3. The new algorithm is globally convergent under some mild assumptions on the objective function. Numerical results are reported in Section 5, which suggest that improvements have been achieved.

## 2. GRADIENT METHODS VIA WEAK SECANT EQUATION

We begin by exploring the properties of the gradient methods for minimizing a strictly convex quadratic function:

$$(2.1) \quad f(x) = \frac{1}{2}x^T Ax - b^T x,$$

where  $A$  is an  $n \times n$  symmetric and positive definite matrix.

Let  $\{x_k\}$  be the sequence generated by (1.2) from an initial  $\{x_0\}$ . By the fact that  $g_k = Ax_k - b$ , it follows that

$$(2.2) \quad g_{k+1} = (I - \alpha_k A)g_k.$$

Without the loss of generality, we assume that  $A$  has distinct eigenvalues

$$0 < \lambda_1 < \lambda_2 < \dots < \lambda_n,$$

and that the components of the gradient  $g_{0,i} \neq 0$  for all  $i = 1, 2, \dots, n$ . By the fact that any gradient method is invariant under any orthogonal transformations and the gradient components corresponding to the identical eigenvalues can be combined (see, for example Fletcher [6]), we can assume that the matrix  $A$  is

$$(2.3) \quad A = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

Denoting  $g_k = (g_{k,1}, g_{k,2}, \dots, g_{k,n})^T$  and by (2.2)-(2.3), we have that

$$(2.4) \quad g_{k+1,i} = (1 - \alpha_k \lambda_i)g_{k,i}.$$

It is clear that from this recurrence that if  $g_{0,i} = 0$  for any  $i$ , then this property will persist for all  $k > 0$ . Moreover if there are any eigenvalues of multiplicity  $m > 1$ , then we can choose the corresponding eigenvectors so that  $g_{0,i} = 0$  for at least  $m - 1$  corresponding indices of  $g_0$  (see Fletcher [6] for details).

Any steplength selection such that  $\|(I - \alpha_k A)\|_2$  is minimized should give a good convergence result by the relation

$$(2.5) \quad \|g_{k+1}\|_2 \leq \|(I - \alpha_k A)\|_2 \|g_k\|_2.$$

One of the choice is the steplength (see [5])

$$(2.6) \quad \alpha_k = \frac{2}{\lambda_1 + \lambda_n}.$$

Since  $\lambda_1$  and  $\lambda_n$  are usually unknown to users, it is normal that this steplength is good only in theory.

Similarly as  $|g_{k+1,i}| = |1 - \alpha_k \lambda_i| |g_{k,i}|$ , it should be good for the sake of convergence if we choose the steplength  $\alpha_k$  such that

$$(2.7) \quad |1 - \alpha_k \lambda_i|, \quad i = 1, 2, \dots, n$$

is minimized.

An obvious choice is letting  $\alpha = 1/\lambda_i$ . Since  $A$  has  $n$  distinct eigenvalues, we will need  $n$  difference scalars as steplengths along each component of the negative gradient direction. Hence, by denoting  $x_k = (x_{k,1}, x_{k,2}, \dots, x_{k,n})^T$ , we have our new gradient method with updating scheme:

$$(2.8) \quad x_{k+1,i} = x_{k,i} - \alpha_{k,i} g_{k,i}$$

where  $\alpha_{k,i}$  is a scalar, which acts as steplength along the  $i$ th component of the negative gradient direction. In shorts, we try to update a new iterate by

$$(2.9) \quad x_{k+1} = x_k - D_k^{-1} g_k$$

where  $D_k = \text{diag}(1/\alpha_{k,1}, 1/\alpha_{k,2}, \dots, 1/\alpha_{k,n})$ .

Equivalent to the problem of minimizing (2.7) with respect to  $\alpha_k$ , it seems reasonable to consider the problem of choosing  $D_k^{-1}$  such that

$$(2.10) \quad \bar{D}_k^{-1} = \arg \min \|(I - D_k^{-1} A)\|_2; \quad D_k^{-1} \text{ is diagonal.}$$

In other words, we seek well-posed metric problems such that the solution  $D_k$  through the diagonal updating is a good approximation of  $A$ . Note that any approximation of  $A$  should satisfy the secant equation  $As_{k-1} = y_{k-1}$  to ensure the curvature information is correct. Furthermore, since  $D_k$  is diagonal (so is  $A$ ) and requires only the storage of an  $n$ -dimensional vector, it is fine to let  $D_k$  satisfies only the weak secant equation:

$$(2.11) \quad s_{k-1}^T D_k s_{k-1} = s_{k-1}^T y_{k-1}.$$

The weak secant relation was first introduced and studied by Dennis and Wolkowicz [4]. Now, it is possible to us to obtain  $D_k$  iteratively as follows:

Suppose that  $D_k > 0$  is a positive definite diagonal matrix and  $D_{k+1}$  is the updated version of  $D_k$  which is also diagonal. Since we require the updated  $D_{k+1}$

to be a valid approximation for  $A$ , then  $D_{k+1}$  must satisfy the weak secant equation and that the deviation between  $D_{k+1}$  and  $D_k$  is minimized under some variational principle, i.e. we seek well-posed metric problems such that the solution  $D_{k+1}$  through the diagonal updating, incorporates available curvature information from the step and the gradient changes as well as that contained in  $D_k$ . As noted earlier, a diagonal matrix simply needs the same computer storage as a vector so an algorithm with  $O(n)$  storage will be maintained. We now state the following theorem which is due to Nazareth et al. [7]:

**Theorem 2.1.** *Let  $\Delta_k = D_{k+1} - D_k$  be the deviation between  $D_{k+1}$  and  $D_k$ ,  $s_k = x_{k+1} - x_k$  and  $y_k = g_{k+1} - g_k$ . Assume that  $s_k \neq 0$  and  $D_k > 0$ . Consider the minimization problem:*

$$(2.12) \quad \begin{aligned} \min \quad & \|\Delta_k\|_F \\ \text{s.t.} \quad & s_k^T \Delta_k s_k = s_k^T y_k - s_k^T D_k s_k \quad \text{and} \quad \Delta_k \text{ is diagonal,} \end{aligned}$$

where  $\|\cdot\|_F$  denotes the Frobenius norm. Then the optimal solution of (2.12) is given by

$$\Delta_{k,i} = \frac{(s_k^T y_k - s_k^T D_k s_k) s_{k,i}^2}{\text{tr}(E^2)}, \quad i = 1, 2, \dots, n$$

where  $\text{tr}$  denotes the trace operator,  $\Delta_{k,i}$  is the  $i$ th diagonal element of  $\Delta_k$ ,  $s_{k,i}$  is the  $i$ th component of  $s_k$  and  $E = \text{diag}(s_{k,1}^2, s_{k,2}^2, \dots, s_{k,n}^2)$ .

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

$$(2.13) \quad d_{k+1,i} = d_{k,i} + \frac{(s_k^T y_k - s_k^T D_k s_k) s_{k,i}^2}{\text{tr}(E^2)}, \quad i = 1, 2, \dots, n$$

where  $d_{k+1,i}$  and  $d_{k,i}$  is the  $i$ th diagonal element of  $D_{k+1}$  and  $D_k$ , respectively. (2.13) is used in the quasi-Cauchy algorithm (a kind of quasi Newton algorithm where the updated matrices are restricted to be diagonal) by Nazareth et al. [7]. However, there is no guarantee that either  $D_{k+1} > 0$  or the gradient method with updating scheme (2.9) is always monotone. Hence some safeguards are needed. The positive-definiteness of  $D_{k+1}$  can be easily verified by checking the condition  $d_{k,i} > 0, \forall i$ . When this is violated, we retain the previous diagonal updating matrix by letting  $D_{k+1} = D_k$ .

In the following we give our monotone strategy:

**Theorem 2.2.** *Assume that  $f(x)$  is a strictly convex quadratic of the form (2.1) and  $\{x_k\}$  be a sequence generated by the updating (2.9) where  $D_k$  is assumed*

to be positive-definite,  $D_{k+1}$  is defined by (2.13) and  $D_0$  is any positive-definite diagonal matrix. Let  $d_{k,m}$ ,  $d_{k,M}$ ,  $d_{k+1,m}$  and  $d_{k+1,M}$  be the smallest and largest diagonal element of  $D_k$  and  $D_{k+1}$ , respectively. If

$$(2.14) \quad d_{k,M}^{-1} - \frac{d_{k,m}^{-2} d_{k+1,m}^{-1}}{2} > 0,$$

and  $\|g_k\| \neq 0$  for all finite  $k$ , then the sequence  $\{f(x_k)\}$  is monotonically decreasing.

*Proof.* Consider the Taylor expansion of (2.1) at  $x_{k+1}$ :

$$(2.15) \quad f(x_{k+1}) = f(x_k - D_k^{-1}g_k) = f(x_k) - g_k^T D_k^{-1}g_k + \frac{1}{2}g_k^T D_k^{-1} A D_k^{-1}g_k.$$

Since  $s_k = x_{k+1} - x_k = -D_k^{-1}g_k$ ,

$$s_k^T A s_k = s_k^T D_{k+1} s_k = g_k^T D_k^{-1} D_{k+1} D_k^{-1} g_k,$$

and

$$d_{k,M}^{-1} \|g_k\|^2 \leq g_k^T D_k^{-1} g_k \leq d_{k,m}^{-1} \|g_k\|^2,$$

(2.15) becomes

$$f(x_{k+1}) \leq f(x_k) - \left( d_{k,M}^{-1} - \frac{d_{k,m}^{-2} d_{k+1,m}^{-1}}{2} \right) \|g_k\|^2.$$

It follows from  $d_{k,M}^{-1} - \frac{d_{k,m}^{-2} d_{k+1,m}^{-1}}{2} > 0$  and  $\|g_k\| \neq 0$  that  $f(x_{k+1}) < f(x_k)$ . ■

### 3. MONOTONE GRADIENT METHOD

We give our algorithm:

#### MONOGRAD Algorithm

- Step 1. Consider an initial point  $x_0$  and a positive-definite diagonal matrix  $D_0$  (we set  $D_0 = I$ ). Compute  $x_1 = x_0 - g_0$  where  $g_0 = \nabla f(x_0)$  and set  $k = 1$ .
- Step 2. Compute  $D_k^{-1} = \text{diag}(1/d_{k,1}, 1/d_{k,2}, \dots, 1/d_{k,n})$  where  $d_{k,i}$ ,  $i = 1, 2, \dots, n$  is given by (2.13).
- Step 3. If  $d_{k-1,M}^{-1} - \frac{d_{k-1,m}^{-2} d_{k,m}^{-1}}{2} > 0$  or  $D_k > 0$  is violated, set  $D_k^{-1} = D_{k-1}^{-1}$ . Else retain  $D_k^{-1}$  that is computed in Step 2.
- Step 4. Update  $x_{k+1} = x_k - D_k^{-1}g_k$ .

Step 5. Test a criterion for stopping the iterations. If the test is satisfied, then stop, else set  $k := k + 1$  and return to Step 2.

Note that MONOGRAD algorithm is different from other monotone gradient methods in such a way that MONOGRAD algorithm requires neither function evaluations nor inexact line search conditions to guarantee monotonicity. Moreover, it is very simple to use. To see just how superior the new method over the BB method is, a simple numerical example is devised. We consider the quadratic function (2.1) where

$$(3.1) \quad n = 50, \quad A = \text{diag}(1, 10, 20, \dots, 10(n - 1)) \quad \text{and} \quad b = (1, 1, \dots, 1)^T.$$

The initial point  $x_0 = (1, 1, \dots, 1)$  as well as the following criterion for stopping the iterations

$$(3.2) \quad \|g_k\| < 10^{-4}$$

are used. The performance of BB and MONOGRAD methods are shown in Fig. 1.

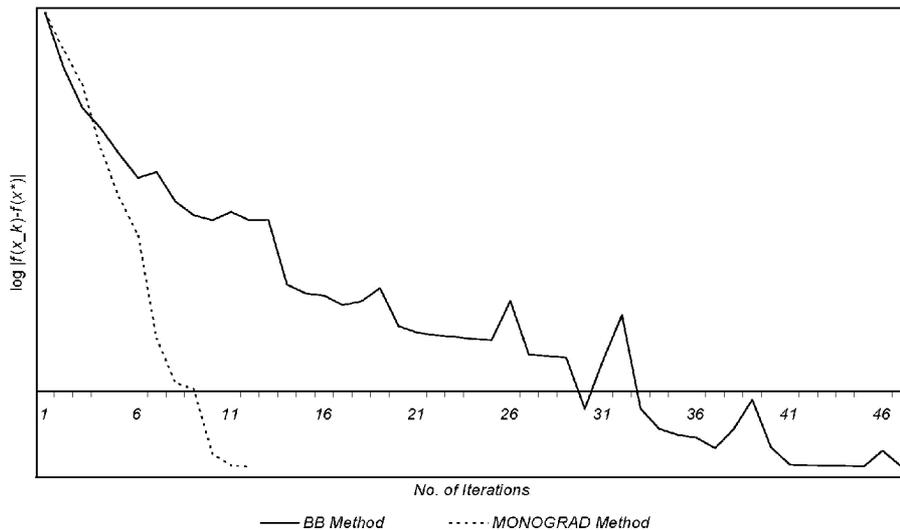


Fig. 1. BB vs MONOGRAD.

Here  $x^*$  denotes the unique minimizer of the objective function and the difference  $f(x_k) - f(x^*)$  is plotted on a log scale against the number of iterations. A noticeable feature of the MONOGRAD method is that monotonically decreasing in function values is observed in all iterations. On the other hand, there are many occasions on which a big jump is seen in  $f(x_k) - f(x^*)$  for the BB method above the slowly varying part of the graph.

## 4. CONVERGENCE ANALYSIS

We shall also establish the convergence of the MONOGRAD algorithm when applied to the minimization of a strictly convex non-quadratic function. For the analysis of this section, we assume that the objective function,  $f$  is strictly convex and bounded below on the closed level set  $S = \{x \in \text{Domain}(f) : f(x) \leq f(x_0)\}$ .

**Theorem 4.3.** *Let  $\{x_k\}$  be a sequence generated by the MONOGRAD Algorithm and  $x^*$  is the unique minimizer of a strictly convex function  $f$ . Then either  $g_k = 0$  holds for some finite  $k \geq 1$  or  $\lim_{k \rightarrow \infty} \|g_k\| = 0$ . Moreover,  $\{x_k\}$  converges  $R$ -linearly to  $x^*$ .*

*Proof.* Denote  $\bar{G} = \nabla^2 f(x_k + \theta s_k)$  for  $\theta \in (0, 1)$ . Again let  $d_{k,m}$ ,  $d_{k,M}$ ,  $d_{k+1,m}$  and  $d_{k+1,M}$  be the smallest and largest diagonal element of  $D_k$  and  $D_{k+1}$ , respectively. Consider the Taylor expansion of the strictly convex function,  $f$  at  $x_{k+1}$ :

$$(4.1) \quad f(x_k - D_k^{-1}g_k) = f(x_k) - g_k^T D_k^{-1}g_k + \frac{1}{2}g_k D_k^{-1} \bar{G} D_k^{-1}g_k.$$

Since  $\bar{G}s_k = y_k$ , it follows that  $s_k^T \bar{G}s_k = g_k D_k^{-1} D_{k+1} D_k^{-1} g_k$ . Thus

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

where  $c = d_{k,M}^{-1} - \left( d_{k,m}^{-2} d_{k+1,m}^{-1} \right) / 2$ . If  $c < 0$  or  $D_{k+1} < 0$ , then we let  $D_{k+1} = D_k$ . This implies that  $s_k^T \bar{G}s_k = g_k D_k^{-1} D_k D_k^{-1} g_k = g_k D_k^{-1} g_k$ . Hence, (4.1) becomes

$$(4.2) \quad f(x_{k+1}) \leq f(x_k) - \frac{1}{2} d_{k,m}^{-1} \|g_{k-1}\|^2.$$

Therefore in both cases, we have that

$$(4.3) \quad f(x_{k+1}) \leq f(x_k) - c \|g_{k-1}\|^2,$$

where  $c > 0$  is either equal to  $d_{k,M}^{-1} - \left( d_{k,m}^{-2} d_{k+1,m}^{-1} \right) / 2$  or  $(2d_{k,m})^{-1}$ .

This implies that  $f(x_{k+1}) \leq f(x_k)$  for all  $k$  and since  $f$  is bounded below, it follows that

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

As  $f(x_k) - f(x_{k+1}) \rightarrow 0$ , and  $c > 0$  then  $\lim_{k \rightarrow \infty} \|g_k\| = 0$ , i.e.  $x_k$  converges to  $x^*$ . Having in mind that  $f(x_k)$  is a nonincreasing sequence, it concludes that  $f(x_k)$  converges to  $f(x^*)$ . Furthermore, the strictly convexity of  $f$  implies that we can bound  $f(x^*)$ :

$$(4.4) \quad f(x) - \frac{1}{2q} \|g(x)\|^2 \leq f(x^*) \leq f(x) - \frac{1}{2Q} \|g(x)\|^2,$$

where  $q$  and  $Q$  are positive constants such that  $q\|z\|^2 \leq z^T \nabla^2 f(x)z \leq Q\|z\|^2$  for all  $x \in S$  and  $z \in R^n$ . The existence of such constants is guaranteed by the convexity of  $f$ . It follows that  $\|g_k\|^2 \geq 2q(f(x_k) - f(x^*))$ . Thus, (4.3) becomes

$$(4.5) \quad f(x_{k+1}) - f(x^*) \leq h(f(x_k) - f(x^*)),$$

where  $h = 1 - cq$ . Note that as  $cq > 0$  and  $f(x_{k+1}) \leq f(x_k)$ , we must have  $0 < h < 1$  for all  $k$ . Therefore the MONOGRAD algorithm is R-linearly convergent.

### 5. NUMERICAL RESULTS

In this section we give some numerical results on BB method and MONOGRAD method for minimization of some convex functions. Our source code is written in Matlab 7.0. Test functions are the standard unconstrained optimization problems that are posted at the website: <http://www.ici.ro/camo/neculai/SCALCG/testuo.pdf> (accessed on Feb '08).

The termination criteria for all methods is

$$(5.1) \quad \|g_k\| < 10^{-5} \times \max\{1, \|x_k\|\}.$$

The numerical comparative results are given in Table 5.1, where the integer numbers are the number of iterations. The symbol – in the table indicates that the method fails to converge within 1000 iterations.

Table 5.1. BB vs MONOGRAD

Test Function (Dimension)	Initial point $x_0$	BB	MONOGRAD
Diagonal2 (10)	$(\frac{3}{1}, \frac{3}{2}, \dots, \frac{3}{n})$	-	50
Diagonal2 (50)	$(\frac{3}{1}, \frac{3}{2}, \dots, \frac{3}{n})$	-	160
Diagonal2 (100)	$(\frac{3}{1}, \frac{3}{2}, \dots, \frac{3}{n})$	-	250
Diagonal2 (500)	$(\frac{3}{1}, \frac{3}{2}, \dots, \frac{3}{n})$	-	582
Perturbed Quadratic (10)	(0.5, 0.5, ..., 0.5)	25	10
Perturbed Quadratic (50)	(0.5, 0.5, ..., 0.5)	61	15
Perturbed Quadratic (80)	(0.5, 0.5, ..., 0.5)	70	19
Perturbed Quadratic (100)	(0.5, 0.5, ..., 0.5)	87	23
Diagonal6 (10)	(1, 2, ..., $n$ )	31	28
Diagonal6 (50)	(1, 2, ..., $n$ )	15	10
Diagonal6 (100)	(1, 2, ..., $n$ )	35	8
Diagonal6 (500)	(1, 2, ..., $n$ )	16	9
Almost Perturbed Quadratic (10)	(0.5, 0.5, ..., 0.5)	17	10
Almost Perturbed Quadratic (50)	(0.5, 0.5, ..., 0.5)	48	14
Almost Perturbed Quadratic (100)	(0.5, 0.5, ..., 0.5)	68	17
Almost Perturbed Quadratic (500)	(0.5, 0.5, ..., 0.5)	196	56

EG2 (20)	(1, 1, ..., 1)	66	16
EG2 (50)	(1, 1, ..., 1)	211	22
EG2 (100)	(1, 1, ..., 1)	59	34
EG2 (500)	(1, 1, ..., 1)	62	51
Quadratic QF1 (10)	(1, 1, ..., 1)	17	8
Quadratic QF1 (50)	(1, 1, ..., 1)	42	13
Quadratic QF1 (100)	(1, 1, ..., 1)	61	16
Quadratic QF1 (500)	(1, 1, ..., 1)	185	27
Raydan2 (10)	(1, 1, ..., 1)	6	5
Raydan2 (50)	(1, 1, ..., 1)	7	5
Raydan2 (100)	(1, 1, ..., 1)	7	5
Raydan2 (500)	(1, 1, ..., 1)	7	5
Quadratic QF2 (10)	(1, 1, ..., 1)	6	6
Quadratic QF2 (50)	(1, 1, ..., 1)	5	5
Quadratic QF2 (100)	(1, 1, ..., 1)	5	5
Quadratic QF2 (500)	(1, 1, ..., 1)	4	4
Diagonal5 (10)	(2, 2, ..., 2)	-	20
Diagonal5 (50)	(2, 2, ..., 2)	-	22
Diagonal5 (100)	(2, 2, ..., 2)	-	22
Diagonal5 (500)	(2, 2, ..., 2)	-	24
Extended Tridiagonal2 (10)	(0.6, 0.6, ..., 0.6)	7	6
Extended Tridiagonal2 (50)	(0.6, 0.6, ..., 0.6)	7	6
Extended Tridiagonal2 (100)	(0.6, 0.6, ..., 0.6)	7	6
Extended Tridiagonal2 (500)	(0.6, 0.6, ..., 0.6)	7	6
Extended Three Exponential Terms (10)	(-0.1, -0.1, ..., -0.1)	14	14
Extended Three Exponential Terms (50)	(-0.1, -0.1, ..., -0.1)	14	14
Extended Three Exponential Terms (100)	(-0.1, -0.1, ..., -0.1)	14	14
Extended Three Exponential Terms (500)	(-0.1, -0.1, ..., -0.1)	14	14
Hager (10)	(2, 2, ..., 2)	-	13
Hager (20)	(2, 2, ..., 2)	19	18
Hager (50)	(2, 2, ..., 2)	23	33

From Table 5.1 we see that MONOGRAD method is superior then the BB method on our set of test problems.

## 6. CONCLUSION

In this paper we proposed a monotone gradient, MONOGRAD method. The proposed method uses a diagonal matrix to represent the "steplengths" in the framework of BB method. Then a monotone strategy is derived and incorporated with the

proposed method. The MONOGRAD method is appealing for several reasons: it is simple to implement, it requires no function calls, low storage requirement, globally converged and possesses linear rate of convergence under very mild conditions.

#### ACKNOWLEDGMENT

This work is partially supported by the Malaysia Fundamental Research Grant Scheme (Grant No: 04-10-07-386FR). The authors would like to thank Prof. Yu-Hong Dai for his helpful comments on the draft of this paper and thanks also due to the anonymous referees, whose comments and suggestions greatly improved this paper.

#### REFERENCES

1. J. Barzilai and J. M. Borwein, Two point step size gradient methods, *IMA J. Numer. Anal.*, **8** (1988), 141-148.
2. Y. H. Dai, J. Y. Yuan and Y. Yuan, Modified two-point stepsize gradient methods for unconstrained optimization, *Comp. Optim. and Appl.*, **22** (2002), 103-109.
3. Y. H. Dai and Y. Yuan, Alternate minimization gradient method, *IMA J. Numer. Anal.*, **23** (2003), 377-393.
4. J. E. Dennis and H. Wolkowicz, Sizing and least change secant methods, *SIAM J. Numer. Anal.*, **30** (1993), 1291-1313.
5. H. C. Elman and G. H. Golub, Inexact and preconditioned Uzawa algorithms for saddle point problems, *SIAM J. Numer. Anal.*, **31** (1994), 1645-1661.
6. R. Fletcher, *On the Barzilai-Borwein method*, Research Report NA/207, University of Dundee, UK, 2001.
7. J. L. Nazareth, H. Wolkowicz and M. Zhu, The quasi-Cauchy relation and diagonal updating, *SIAM J. Optim.*, **9** (1999), 1192-1204.
8. M. Raydan, On the Barzilai and Borwein choice of steplength for the gradient method, *IMA J. Numer. Anal.*, **13** (1993), 618-622.
9. Y. Yuan, A new stepsize for the steepest descent method, *J. Comp. Math.*, **24** (2006), 149-156.

Wah June Leong, Malik Abu Hassan and Mahboubeh Farid  
Institute for Mathematical Research,  
University Putra Malaysia,  
43400 Serdang,  
Selangor, Malaysia  
E-mail: leong@math.upm.edu.my  
malik@fsas.upm.edu.my  
mahboubeh\_farid@yahoo.com