New step sizes of the gradient methods for unconstrained optimization problem

Basim A. Hassan
Department of Mathematics
College of Computers Sciences and Mathematics
University of Mosul
Mosul
Iraq
basimah@uomosul.edu.iq

Fadhil Alfarag*
School of Mathematics
Birmingham University
United Kingdom
fadhil68@hotmail.co.uk

Snezana Djordjevic
University of Nis
Faculty of Technology
Leskovac
Serbia
snezanadjordjevicle@gmail.com

Abstract. In this work, we derive a new formula for step-length in the frame of gradient descent methods. The idea of method is based on the quadratic model for using a new approximation of the Hessian of the minimizing function. The rate of convergence is linear and it belongs to the same class of gradient descent methods. Numerical experiment shows that new method is very promising compared to the Barzilai-Borwein approach.

Keywords: modified Newton's method, iteration methods.

1. Introduction

Unconstrained optimization methods are useful in finding the optimal solution for continuous and differentiable functions.

An unconstrained optimization problems can be stated as follows:

$$\min \Psi(x), x \in R^n. \tag{1}$$

The method of steepest descent is generally the most simplest method for unconstrained optimization problem. More details can be found in[10].

∗. Corresponding author

In the steepest descent method, at each iterate point x_k, the search direction is chosen as the steepest descent direction:

$$d_k = -\nabla\Psi(x_k). \tag{2}$$

The iterative procedure of this method can be stated as follows:

$$x_{k+1} = x_k - \alpha_k \nabla\Psi(x_k). \tag{3}$$

where α_k is step-size, to be computed by line search procedure along the search direction d_k. For more details see [9].

Now, we remind to several modified gradient methods with some modified step-sizes, due to its simplicity and numerical efficiency.

One of the famous methods is BB methods [5] suggested by Barzilai and Borwein in 1988, the step-sizes is given by:

$$\alpha_k^{BB1} = \frac{\|s_k\|^2}{y_k^T s_k}\ ,\ \alpha_k^{BB2} = \frac{y_k^T s_k}{\|y_k\|^2}, \tag{4}$$

where $s_k = x_{k+1} - x_k = -\alpha_k \nabla\Psi(x_{k+1})$ and $y_k = \nabla\Psi(x_{k+1}) - \nabla\Psi(x_k)$.

In 2002, Dai et al. [7] developed some new step-sizes for BB-like methods, given by:

$$\alpha_k^{DYY1} = \frac{s_k^T s_k}{2(\Psi_k - \Psi_{k+1} + g_{k+1}^T s_k)}\ ,\ \alpha_k^{DYY2} = \frac{s_k^T s_k}{6(\Psi_k - \Psi_{k+1}) + 4g_{k+1}^T 2s_k} \tag{5}$$

However, gradient methods are worthy in terms of the numerical performances. See also [2,6,8,11,12,13] for more details.

It remains to examine how to design more effective and efficient gradient methods for unrestricted improvement. Some new step sizes are described and tested.

2. Deriving new step sizes

We derive some new optimal step sizes of gradient methods based on the second-order Taylor's series approximation of the function. The model function is defined by:

$$\Psi(x_k) = \Psi(x_{k+1}) - g_{k+1}^T s_k + \frac{1}{2} s_k^T Q s_k. \tag{6}$$

Obviously, the approximate optimal step size associated to $\Psi(x_k)$ is:

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T Q d_k}. \tag{7}$$

Let $\tau_k I$ be an approximation to the Hessian. Now, from the exact search property the following condition must be fulfilled:

$$g_{k+1}^T s_k = 0. \tag{8}$$

It is clear that the selection of τ_k is key to the approximate effect of the Hessian matrix. We determine τ_k by putting (8) in (6) we get:

$$2(\Psi(x_k) - \Psi(x_{k+1})) = \tau_k s_k^T s_k. \tag{9}$$

which implies that:

$$\tau_k = \frac{2(\Psi(x_k) - \Psi(x_{k+1}))}{s_k^T s_k} \tag{10}$$

We obtain the approximation optimal step-size:

$$\alpha_k = \tau_k^{-1} = \frac{s_k^T s_k}{2(\Psi(x_k) - \Psi(x_{k+1})}. \tag{11}$$

Imposing $g_{k+1}^T d_k = 0$ and using (7) which implies that:

$$s_k^T Q s = 2\Psi(x_k) - \Psi(x_{k+1}) - \alpha_k g_k^T d_k/2 \tag{12}$$

which also implies a new approximation model:

$$\tau_k s_k^T s_k = \Psi(x_k) - \Psi(x_{k+1}) - \alpha_k g_k^T d_k/2 \tag{13}$$

which implies that:

$$\tau_k = \frac{\Psi(x_k) - \Psi(x_{k+1}) - g_k^T s_k/2}{s_k^T s_k}. \tag{14}$$

Similarly, we obtain the approximation optimal step-size:

$$\alpha_k = \tau_k^{-1} = \frac{s_k^T s_k}{\Psi(x_k) - \Psi(x_{k+1} - g_k^T s_k/2}. \tag{15}$$

As a result, we adopt a new algorithm and called Algorithm BFS.

Algorithm 1 Algorithm BFS

1: Select $x_o \in R$ and compute $d_k = -\nabla\Psi(x_k)$. Set $k = 0$.
2: Test a criterion for stopping the iterations. If the test is satisfied, then stop,
3: If $\Psi(x_k) - \Psi(x_{k+1}) < 0.5$, then put $\Psi(x_k) - \Psi(x_{k+1}) = 0.5$ and compute α_k the scalars as in (11) and (15), using these scalars.
4: Update the variables: $x_{k+1} = x_k - \alpha_k \nabla\Psi(x_k)$ set and go to step 2.

3. Global convergence

Now we turn to the global convergence of the BFS under the following assumption:

Let Ψ is strongly convex function and let the level set $\Phi = \{x \in R^n : \Psi(x) \leq \Psi(x_o)\}$ is closed. Strong convexity of Ψ on Φ involves the existence of the constant m and M such that:

$$mI \leq \nabla^2\Psi(x) \leq MI, \tag{16}$$

for all $x \in \Phi$. A consequence of strong convexity of Ψ on Φ is that we can bound Ψ^* as:

$$\Psi(x) - \frac{1}{2m}\|\nabla^2\Psi(x)\|_2^2 \leq \Psi(x^*) \leq \Psi(x) - \frac{1}{2M}\|\nabla^2\Psi(x)\|_2^2. \tag{17}$$

More details, can be found in [5,6].

Theorem 3.1. *The New Algorithm with backtracking is linearly convergent and*

$$\Psi(x_k) - \Psi^* \leq \left\{\prod_{i=1}^{k-1} c_i\right\}\{\Psi(x_o) - \Psi^*\}, \tag{18}$$

where $c_1 = 1 - \min\{m, ms^{p_k}\} < 1$ *and* $p_k \geq 1$ *is an integer,* ($p_1 = 1, 2, 3, \cdots$ *given by the backtracking procedure*).

Proof. By (10), we write:

$$\Psi(x_{k+1}) = \Psi(x_k) - 1/2(\tau_k s_k^T s_k). \tag{19}$$

Now, at the point $x_{k+1} = x_k - \alpha_k\nabla\Psi(x_k)$ we have:

$$\begin{aligned} \Psi(x_{k+1}) &= \Psi(x_k) - 1/2(\tau_k\alpha_k^2\nabla\Psi(x_k)^T\nabla\Psi(x_k)), \\ \Psi(x_{k+1}) &= \Psi(x_k) - 1/2(\alpha_k\nabla\Psi(x_k)^T\nabla\Psi(x_k)). \end{aligned} \tag{20}$$

Hence

$$\Psi(x_{k+1}) = \Psi(x_k) - \frac{1}{2}\alpha_k\|\nabla\Psi(x_k)\|_2^2. \tag{21}$$

Using backtracking procedure terminates either with $\alpha_k = 1$ or with $\alpha_k = s^{p_k}$, where p_k is an integer. Therefore:

$$\Psi(x_{k+1}) = \Psi(x_k) - \min\{\frac{1}{2}, \frac{1}{2}s^{p_k}\}\|\nabla\Psi(x_k)\|_2^2. \tag{22}$$

Having in view that for strongly convex functions $\|\nabla\Psi(x_k)\|_2^2 \geq 2m(\Psi(x_k) - \Psi^*)$ it follows that:

$$\Psi(x_k) - \Psi^* \leq c_k(\Psi(x_k) - \Psi^*), \tag{23}$$

where $c_k = 1 - \min\{m, ms^{p_k}\}$.

On the other hand, from (15), we get:

$$\tau_k s_k^T s_k = \Psi(x_k) - \Psi(x_{k+1}) - \alpha_k g_k^T d_k/2. \tag{24}$$

Also, at the point $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$ and $d_k = -\nabla\Psi(x_k)$ we have:

$$\tau_k \alpha_k^2 \nabla\Psi(x_k)^T \nabla\Psi(x_k) = \Psi(x_k) - \Psi(x_{k+1}) - \alpha_\nabla \Psi(x_k)^T \nabla\Psi(x_k)/2. \tag{25}$$

From the above equation, we get:

$$\begin{gathered} \alpha_k \|\nabla\Psi(x_k)\|_2^2 = \Psi(x_k) - \Psi(x_{k+1}) + \alpha_k \|\nabla\Psi(x_k)\|_2^2/2, \\ \Psi(x_k) - \Psi(x_{k+1}) = (\alpha_k/2)\|\nabla\Psi(x_k)\|_2^2, \end{gathered} \tag{26}$$

which implies that:

$$\Psi(x_{k+1}) = \Psi(x_k) - \frac{1}{2}\alpha_k \|\nabla\Psi(x_k)\|_2^2. \tag{27}$$

Similarly in case above, we obtain:

$$\Psi(x_{k+1}) - \Psi^* \leq c_k(\Psi(x_k) - \Psi^*), \tag{28}$$

where $c_k = 1 - \min\{m, ms^{p_k}\}$. Since $c_k < 1$ the sequence $\{\Psi(x_k)\}$ is linearly convergent, as a geometric series, to Ψ^*. □

4. Numerical results

We implemented the BFS and BB methods to a great number of standard test functions include unconstrained problems which are mainly from [1], to check the numerical performance. Different test functions have been used in different research such as [3, 4].

The iteration is stopped if the inequality $\| \Psi(x) \| \leq 10^{-6}$ or $\frac{\Psi(x_{k+1}) - \Psi(x)}{1 + \Psi(x_k} \leq 10^{-6}$ is satisfied. We choose the following parameters for BFS and BB: $\alpha = 0.00001$ and $s = 0.8$, then the number of iterations (NI) and the the number of function evaluations (NF) corresponding to the BFS and Barzilai- Borwein algorithms are given in Table 1.

		BB algorithm		BFS with (11)		BFS with (15)	
P. No.	**n**	**NI**	**NF**	**NI**	**NF**	**NI**	**NF**
Extended Tridiagonal 1	100	45	215	16	98	7	28
	1000	45	215	20	136	7	28
Generalized Tridiagonal 2	100	36	124	28	210	31	135
	1000	90	516	44	365	24	77
Generalized PSC1	100	794	4808	14	103	12	49
	1000	803	4745	30	306	21	50
Extended BD1	100	21	69	20	135	20	73
	1000	21	69	27	182	25	104
Quadratic QF1	100	149	649	30	208	31	132
	1000	1364	7145	94	695	90	372
Extended Quadratic Penalty QP1	100	13	58	20	173	11	52
	1000	17	73	15	126	6	49
Extended Tridiagonal 2	100	42	146	9	41	10	34
	1000	37	125	13	84	14	52
DQDRTIC (CUTE)	100	144	717	41	309	42	178
	1000	175	947	44	431	44	193
DIXMAANC (CUTE)	100	9	33	18	142	10	35
	1000	10	35	26	244	30	41
DIXMAANE (CUTE)	100	70	276	28	209	11	32
	1000	335	1618	46	407	16	42
Broyden Tridiagonal	100	132	545	23	179	33	137
	1000	104	569	44	304	40	175
Almost Perturbed Quadratic	100	153	661	25	186	28	124
	1000	1463	7534	75	608	72	303
Tridiagonal Perturbed Quadratic	100	179	813	22	163	28	131
	1000	1347	7030	69	543	72	297
ENGVAL1 (CUTE)	100	30	100	25	198	19	71
	1000	1131	7616	906	6916	27	93
Extended Wood	100	1097	7416	212	1806	21	81
	1000	1131	7616	906	6916	27	93
Total		**9888**	**54968**	**2017**	**15793**	**789**	**3219**

Table 1: Comparing different conjugate gradient methods with different test functions.

5. Conclusion

This paper presents a new deriving method for stepsizes selection by using quadratic model function and considered are modified of the Barzilai and Borwein stepsizes. Based on the numerical results show that the the new algorithms compares good and constructive with the Barzilai-Borwein approach.

Acknowledgement

The authors are very grateful to the University of Mosul / College of Computers Sciences and Mathematics for their provided facilities, which helped to improve the quality of this work.

References

[1] N. Andrie, *An unconstrained optimization test functions collection*, Advanced Modeling and Optimization, 10 (2008), 147-161.

[2] N. Andrei, N., *A new gradient descent method with anticipative scalar approximation of hessian for unconstrained optimization, research institute for informatics*, Center for Advanced Modeling and Optimization, 2005, 1-10.

[3] A. Hassan Basim, *A globally convergence spectral conjugate gradient method for solving unconstrained optimization problems*, Raf. J. of Comp. & Math's., 10 (2013), 21-28.

[4] A. Hassan Basim, *Development a special conjugate gradient algorithm for solving unconstrained minimization problems*, Raf. J. of Comp. & Math's., 9 (2012), 73-84.

[5] J. Barzilai, J.M. Borwein, *Two-point step size gradient methods,* IMA J. Numer. Anal., 8 (1988), 141–148.

[6] F. Biglari, M. Solimanpur, *Scaling on the spectral gradient method,* J. Optim. Theory Appl., 158 (2013), 626–635.

[7] Y.H. Dai, J.Y. Yuan, Y.X. Yuan, *Modified two-point stepsize gradient methods for unconstrained optimization,* Comput. Optim. Appl., 22 (2002), 103–109.

[8] L. Hongwei, L. Zexian, D. Xiaoliang, *A new adaptive Barzilai and Borwein method for unconstrained optimization*, Optim Lett., 2017.

[9] F. Prieto, J. Moggaerza, *Combining search directions using gradient flows*, Mathematical Programming, 96 (2003), 529-559.

[10] S.S. Rao, *Optimization theory and applications*, Wiley Eastern Limited, 2009.

[11] L. Zexian, L. Hongwei, *An efficient gradient method with approximate optimal stepsize for large-scale unconstrained optimization*, Numerical Algorithm, 2017.

[12] Liu Zexian Liu, Liu Hongwei Liu, *Several efficient gradient methods with approximate optimal stepsizes for large scale unconstrained optimization*, Journal of Computational and Applied Mathematics, 328 (2018), 400–413.

[13] L. Zexian, L. Hongwei, H. Chuanmei, Y. Jun, L. Ming, *New scaling on the gradient method*, Journal of Physics: Conf. Series, 2019.

Accepted: April 04, 2020