A New Type of Step Sizes for Unconstrained Optimization

Basim A. Hassan¹*, Maha S. Younis², Mohammed W. Taha³ and Abdulkarim Hassan Ibrahim⁴

¹ Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, IRAQ
² Department of Mathematics, College of Education for Pure Science, University of Mosul, IRAQ
³ Ministry of Education of Iraq, Apartment of Nineveh, IRAQ
⁴Department of Mathematics, Faculty of Science, King Mongkut’s University of Technology Thonburi (KMUTT), Bangkok, Thailand

E-mail: basimah@uomensol.edu.iq

Abstract. Step sizes is very important for a global convergence gradient method for solving the problems of unconstrained optimization. The new step sizes formulas techniques proposed, the key idea used in the construction of the algorithm is to approximate Hessian by a suitable diagonal matrix, which has been found to be the most efficient in this paper. Under weaker conditions, we define the convergences of the proposed methods. In addition, we will show that performance of proposed algorithm is better than of the gradient descent (GD) method.

1. INTRODUCTION

The methods of gradient method are useful in finding the optimum solution of smooth functions. The gradient method problems can be stated as follows:

\[ \text{Min } f(x), \ x \in \mathbb{R}^n \] (1)

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \), whose gradient \( \nabla f(x) \), and the gradient of objective function is denoted by \( g(x) \). For more details see [1].

Widely that the steepest descent direction [2], which looks like this:

\[ d_k = - \nabla f(x_k) \] (2)

In order to solve gradient method problems (1), the following iterative is proved to be an effective approach:

\[ x_{k+1} = x_k - \alpha_k \nabla f(x_k) \] (3)

where \( x_k \) is the current iteration point, \( \alpha_k \) a step size in some line search. For more details see [3].

In the literature, much effort has been devoted to developments new step-sizes methods, due to its simplicity and computational efficiency. Let us elucidate some of the developments the step-sizes wellknown as follows.

For the step-length, Barzilai and Borwein [4] suggested interesting two options:

\[ \alpha_k^{BB 1} = \left\| s_k \right\| \left\| y_k \right\|, \quad \alpha_k^{BB 2} = \frac{y_k^T s_k}{\left\| y_k \right\|^2} \] (4)
where \( s_k = x_{k+1} - x_k \) and \( y_k = g_{k+1} - g_k \).

In 2002, Dai et al. [5] developed some new step-sizes for BB-like methods, given by:

\[
\alpha_k^{(1)} = \frac{s_k^T s_k}{2(\beta_k - s_k^T g_k)}, \quad \alpha_k^{(2)} = \frac{\alpha_k^{(1)} s_k}{6(\beta_k - s_k^T g_k) + 4g_k^T s_k + 2s_k^T s_k}
\]

Moreover, numerical results indicate that gradient methods are worthy and performs much better than the steepest descent method. See also [6,7,8,9,10,11] for more details.

Development some efficient approximate optimal step sizes to design more robust gradient methods and study their important properties.

2. DERIVATION OF THE NEW STEP SIZES

Can be used to improve the traditional gradient methods efficiently by using quadratic model. The details deriving are as follows. First, quadratic model for the objective function \( f(x) \) is as following:

\[
f_{k+1} = f_k + s_k^T g_k + \frac{1}{2} s_k^T Q(x_k) s_k
\]

We define that notation \( a \times I \) is an approximation of the matrix \( Q \) we get:

\[
f_{k+1} = f_k + s_k^T g_k + \frac{1}{2} a s_k^T s_k
\]

We know the minimum of \( s_k \) occurs at:

Let \( \nabla f_{k+1} = 0 \) then \( s_k = -\frac{g_k}{a} \)

Using this knowledge we solve for \( a \) we get:

\[
f_{k+1} = f_k - \frac{1}{a} g_k g_k + \frac{1}{2a^2} a g_k^T g_k
\]

\[
= f_k - \frac{2}{2a} g_k^T g_k + \frac{1}{2a} g_k^T g_k
\]

Thus, we have:

\[
f_{k+1} - f_k = -\frac{1}{2a} g_k^T g_k \quad \text{and} \quad f_{k+1} - f = -\frac{1}{a} g_k^T g_k - \frac{1}{2} g_k^T s_k
\]

resulting in:

\[
a = -\frac{g_k^T g_k}{2(f_{k+1} - f_k)} \quad \text{and} \quad a = -\frac{g_k^T g_k}{1/2 g_k^T s_k + (f_{k+1} - f_k)}
\]

By substituting (11) back into (8), we know use this to solve for \( s_k \) to obtain:

\[
s_k = 2 \frac{(f_{k+1} - f_k)}{g_k^T g_k} g_k \quad \text{and} \quad s_k = \frac{1/2 g_k^T s_k + (f_{k+1} - f_k)}{g_k^T g_k} g_k
\]

It follows from the meaning of \( s_k = \alpha_k d_k \), we get:

\[
\alpha_k d_k = 2 \frac{(f_{k+1} - f_k)}{g_k^T g_k} g_k \quad \text{and} \quad \alpha_k d_k = \frac{1/2 g_k^T s_k + (f_{k+1} - f_k)}{g_k^T g_k} g_k
\]

Multiplying equations (13) by \( g_k^T \), we obtained:

\[
\alpha_k g_k^T d_k = 2 (f_{k+1} - f_k) \quad \text{and} \quad \alpha_k g_k^T d_k = 1/2 g_k^T s_k + (f_{k+1} - f_k)
\]

Yielding:

\[
\alpha_k = \frac{2(f_{k+1} - f_k)}{g_k^T d_k} \quad \text{and} \quad \alpha_k = \frac{1/2 g_k^T s_k + (f_{k+1} - f_k)}{g_k^T d_k}
\]

The new step sizes formulas, called BKA1 and BKA2, respectively.

As a outcome, we take on a new algorithm and called Algorithms BKA.
New Algorithm (Algorithm BKA)

Stage 1. Select \( x_0 \in R^n \) and compute \( d_0 = -g_0 \). Set \( k = 0 \).

Stage 2. If convergence criterion is satisfied, then stop.

Stage 3. Compute the scalars \( \alpha_k \) as in (15), using these scalars.

Stage 4. Update the variables: \( x_{k+1} = x_k - \alpha_k g_k \), set and go to step 2.

3. GLOBAL CONVERGENCE

In order for our new algorithm to achieve global convergence, we must also make the following assumptions:

Let \( f \) is strongly convex function and let the level set \( \mathcal{E} = \{ x \in R^n : f(x) \leq f(x_0) \} \) is closed.

Strong convexity of \( f \) on \( \mathcal{E} \) involves the existence of the constant \( m \) and \( M \) such that:

\[
m I \leq \nabla^2 f(x) \leq M I
\]

for all \( x \in \mathcal{E} \). A consequence of strong convexity of \( f \) on \( \mathcal{E} \) is that we can bound \( f' \) as:

\[
f(x) - \frac{1}{2m} \left\| \nabla^2 f(x) \right\|_2^2 \leq f(x') \leq f(x) - \frac{1}{2M} \left\| \nabla^2 f(x) \right\|_2^2
\]

More details can be found in [6].

Theorem 1.

The New Algorithm with backtracking is linearly convergent and

\[
f(x_k) - f^* \leq \left( \prod_{i=0}^{k-1} c_i \right) (f(x_0) - f^*)
\]

where \( c_i = 1 - \min \{ m, m s^p_i \} \) \( a \) \( p \) is an integer, \( ( p_k = 1, 2, 3, \ldots ) \) given by the backtracking procedure.

Proof:

Using (14), we write:

\[
f(x_{k+1}) = f(x_k) + 1/2 (g_k^T s_k)
\]

Since \( x_{k+1} = x_k - \alpha_k g_k \) we obtain:

\[
f(x_{k+1}) = f(x_k) - 1/2 (\alpha_k \left\| g_k \right\|_2^2)
\]

Using backtracking procedure terminates either with \( \alpha_k = 1 \) or with \( \alpha_k = s^p \) where \( p_k \) is an integer. Therefore:

\[
f(x_{k+1}) = f(x_k) - \min \left\{ \frac{1}{2}, \frac{1}{2} s^p \right\} \left\| g_k \right\|_2^2
\]

Having in view that for strongly convex functions \( \left\| g_k \right\|_2^2 \geq 2m (f(x_k) - f^*) \) it follows that:

\[
f(x_{k+1}) - f^* \leq c_k (f(x_k) - f^*)
\]

On the other hand, again from (14), we get:

\[
\alpha_k g_k^T d_k = 1/2 g_k^T s_k + f(x_{k+1}) - f(x_k)
\]

Again using \( x_{k+1} = x_k - \alpha_k g_k \) and \( d_k = -g_k \) we obtain:

\[
-\alpha_k g_k^T g_k = -1/2 \alpha_k g_k^T g_k + f(x_{k+1}) = f(x_k)
\]

From (24) we get:

\[
f(x_k) - f(x_{k+1}) = 1/2 \alpha_k \left\| g_k \right\|_2^2
\]

Implies:

\[
f(x_{k+1}) = f(x_k) - 1/2 \alpha_k \left\| g_k \right\|_2^2
\]

Similarly in case above, we obtain:
\[ f(x_{k+1}) - f^* \leq c_i (f(x_k) - f^*) \]
where \( c_i = \min \{ m, m s^n \} \). Since \( c_i < 1 \) the sequence \( \{ f(x_k) \} \) is linearly convergent, as a geometric series, to \( f^* \).

3. NUMERICAL RESORTS

We implemented the new and gradient descent algorithm (GD) methods the unconstrained problems, which are often from [12], to check the numerical performance. Some different test functions are given [13,14].

If the inequality is observed, the iteration is terminated \( \| \nabla f(x) \| \leq 10^{-6} \) or \( \alpha_i |g_r^T d_i| \leq 10^{-20} |f_i| \) is satisfied. For new and gradient descent (GD) algorithms, the following parameters were chosen: \( \alpha = 0.0001 \) and \( s = 0.8 \), then the number of iterations (NI) and the number of function evaluations (NF) corresponding to the new and gradient descent (GD) algorithms are given in Table 1. Dolan and Moré [15] presented an appropriate technique to demonstrate the performance of the profiles, which is a statistical process. Other idea have been used for improved gradient methods in various researches such as [16-26].

We can see from Figure 1 and 2 show that BKA1 and BKA2 methods successfully reach the solution points and have performed very good.

![Figure 1: Performance profiles with respect to the number of iterations](image1)

![Figure 2: Performance profiles with respect to the number of function evaluations](image2)
4. CONCLUSIONS

In this paper, we derive a new step size whose idea is based on the quadratic model. As well as the global convergence of the method is analyzed. Numerical outcomes by employing a set of large-scale test problems indicated that BKA 1 and 2 are highly efficient compared to the gradient descent (GD) method.

REFERENCES

[1] Basim A. Hassan and Omer M. E., (2014), A New sufficient descent Conjugate Gradient Method for Nonlinear Optimization, Iraqi Journal of Statistical Sciences, (26), pp.12-24.
[2] Rao, S.S. (2009). Optimization Theory and Applications. Wiley Eastern Limited.
[3] Prieto, F. and Moggaerza, J. (2003). Combining Search Directions using Gradient flows. Mathematical Programming 96, 529-559.
[4] Barzilai, J., Borwein, J.M.: Two-point step size gradient methods. IMA J. Numer. Anal. 8, 141–148 (1988).
[5] Dai, Y.H., Yuan, J.Y., Yuan, Y.X., (2002), Modified two-point stepsize gradient methods for unconstrained optimization. Comput. Optim. Appl. 22, 103–109.
[6] Andrei, N., (2005), A new Gradient Descent Method with Anticipative Scalar Approximation of Hessian for Unconstrained Optimization, Research institute for informatics, Center for Advanced Modeling and Optimization, pp. 1-10.
[7] Biglari, F., Solimanpur, M., (2013), : Scaling on the spectral gradient method. J. Optim. Theory Appl. 158(2), pp.626–635.
[8] Hongwei L., Zexian L. and Xiaoliang D., (2017), A new adaptive Barzilal and Borwein method for unconstrained optimization, Optim. Lett.
[9] Zexian L. and Hongwei L., (2017), An efficient gradient method with approximate optimal stepsizes for large-scale unconstrained optimization, Journal of Computational and Applied Mathematics, 328, pp. 400–413.
[10] Zexian L., Hongwei L., Chuanmei H., Jun Y. and Ming L., (2019), New scaling on the gradient method, Journal of Physics: Conf. Series.
[11] Zexian L., Hongwei L., Chuanmei H., Jun Y. and Ming L., (2019), New scaling on the gradient method, Journal of Physics: Conf. Series.
[12] Andrie N. (2008) ' An Unconstrained Optimization Test functions collection ' Advanced Modeling and optimization. 10, pp.147-161.
[13] Abbas Y. A. and Basim A. H., (2011), A Spectral Conjugate Gradient Method with Inexact line searches, Iraqi Journal of Statistical Science (20), pp.155-163.
[14] Abbas Y. A. and Basim A. H., (2011), A New Type of Conjugate Gradient Method with a Sufficient Descent Property, Iraqi Journal of Statistical Science (20), pp.192-201.
[15] Dolan, E.D., Moré, J.J. (2020), Benchmarking optimization software with performance profiles. Math. Program , 91(2), PP.201–213.
[16] Sun, M., Wang, X., Feng, D. (2017), A family of conjugate gradient methods for large-scale nonlinear equations. Journal of Inequalities and Applications 236, 1–8.
[17] Arazm, M.R.; Babaie-Kafaki, S.; Ghanbari, R. (2017)An extended Dai–Liao conjugate gradient method with global convergence for nonconvex functions. Glas. Mat. 52(72), 361–375.
[18] Malik M., Mamat M., Abas S. S., , (2020), Convergence analysis of a new coefficient conjugate gradient method under exact line search, International Journal of Advanced Science and Technology, 29(5), 187–198.
[19] Malik M., Abas S. S., Mamt M., Sukono, Mohammed I. S., (2020), A new hybrid conjugate gradient method with global convergence properties, International Journal of Advanced Science and Technology, 29(5), 199–210.
[20] Zheng X. and Shi J., (2018), modified sufficient descent polak–riëbiére–polyak type conjugate gradient method for unconstrained optimization problems. Algorithm, 1-10.
[21] Sulaiman I. M., Yakubu U. A. and Mamat M., (2020) Application of Spectral conjugate gradient Method for Solving Unconstrained Optimization Problems. An International Journal of Optimization and Control: Theories & Application, 10(2), 198-205.

[22] Aminifard Z. and Babaie-Kafaki S., (2019), A Modified descent Polak–Ribiére–Polyak conjugate gradient method with global convergence property for nonconvex functions. Calcolo, 56(16), 1-11.

[23] Babaie-Kafaki S. and Ghanbari R., (2017) an optimal extension of the Polak-Ribière-Polyak conjugate gradient method. Numer. Funct. Anal. Optim, 38(9), 1115–1124.

[24] Basim A. Hassan, (2019), A new formula for conjugate parameter computation based on the quadratic model, Indonesian Journal of Electrical Engineering and Computer Science, 3, pp. 954-961.

[25] Basim A. H. Hussein O. D. and Azzam S. Y. A new kind of parameter conjugate gradient for unconstrained optimization, Indonesian Journal of Electrical Engineering and Computer Science, 2020; 17: 404-411.

[26] Basim A. Hassan, Osama M.T. W. and Ayad A. M. A Class of Descent Conjugate Gradient Methods for Solving Optimization Problems, HIKARI Ltd, Applied Mathematical Sciences, 2019; 12, 559 – 567.