Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 27 No. 2 (2024)

Novel Step Size for Unconstrained Optimization

Yoksal A. Laylani[1], Basim A. Hassan[2], Hakeem N. Hussein, Mohammed W. Taha[3], Hawraz N. Jabbar[1]

[1] Department of Mathematic, College of Sciences, Kirkuk University, IRAQ

[2]Department of Mathematic, College of Computer Sciences and Mathematic, Mosul University, IRAQ

[3] Ministry of Education of Iraq, Apartment of Nineveh, IRAQ

Article History:

Received: *24-01-2024*

Revised: *22-03-2024*

Accepted: *21-04-2024*

Abstract: The step length size is the basis of gradient descent method with the desirable conjugate property. The aim of this document is to present a fresh gradient descent technique that calculates the step size based on a basic estimation of the Hessian of the function being minimized. The algorithm that results falls within the category of gradient descent methods with linear convergence characteristics. Our study includes numerical data that demonstrates the significant advantages of our approach compared to traditional gradient descent methods.

Keywords: Unconstrained Optimization, New Kind of Step Sizes, Numerical results.

1. Introduction

Many different approaches have been suggested in academic writing to solve the challenging issue of large scale unconstrained problems:

$$\min\{f(x): |x \in R^n\} \qquad (1)$$

Such that $f: R^n \to R$, R^n is called the n-dimensional Euclidean space. (see e.g. [10,13]).

Gradient descent methods are a crucial set of techniques used to solve (1), particularly for issues that exhibit the following pattern:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) \qquad (2)$$

Such that x_κ is the current iterate, α_k is a positive constant denoted as the step_length and it can be calculate as:

$$\alpha_k = -g_k^T d_k / d_k^T Q_{k+1} d_k, \qquad (3)$$

by some of the line_search, and d_k is the_direction of search generated by the rule:

$$d_k = -\nabla f(x_k) \qquad (4)$$

see [12] for details.

Several suggestions have been made for adapting gradient calculation by using altered step lengths, as this method is known for its simplicity and computational effectiveness. The one of the first method, and the simplest one, noted by Barzilai and Borwein 1988, the step-length was given as:

$$\alpha_k^{BB1} = \frac{\|s_k\|^2}{y_k^T s_k}, \alpha_k^{BB2} = \frac{y_k^T s_k}{\|y_k\|^2} \quad (5)$$

where $s_\kappa = x_{\kappa+1} - x_\kappa = \alpha_\kappa \nabla f(x_\kappa)$ and $y_k = \nabla f(x_{\kappa+1}) - \nabla f(x_\kappa)$. More details can be found in [6].

In 2002, Dai et al. [8] suggested a gradient descent method which takes a different strategy to choose the step size, given by :

$$\alpha_k^{DYY1} = \frac{s_k^T s_k}{2(f_k - f_{k+1} + g_{k+1}^T s_k)}, \alpha_k^{DYY2} = \frac{s_k^T s_k}{6(f_k - f_{k+1}) + 4g_{k+1}^T s_k + 2g_k^T s_k} \quad (6)$$

Based on quadratic model, Basim et al. [5] we developed two efficient approximate optimal stepsizes α_k was updated by the formulas:

$$\alpha_k = \frac{2(f_{k+1} - f_k)}{g_k^T d_k}, \alpha_k = \frac{1/2 g_k^T s_k + (f_{k+1} - f_k)}{g_k^T d_k} \quad (7)$$

The efficiency of all these gradient methods significantly depends on the using suitable values of the stepsizes α_k.

Some other known methods based on the gradient descent strategy to solve the large-scale problems may be found in the literature [1,7,11,14,15,16].

The aim of paper is to offer a new gradient_descent algorithms in which a step length is calculated by backtracking using a simple Hessian approximation based on the 2nd Taylor's series and study their important property. Numerical results show improvement in performance.

2. **Deriving A New Kind Step Size**

New optimal phase sizes for gradient methods are revealed by examining the function through the 2nd Taylor's series approximation. The 2nd Taylor's series can be described as:

$$f(x) = f(x_{\kappa+1}) + g_{\kappa+1}^T (x - x_{\kappa+1}) + \frac{1}{2}(x - x_{\kappa+1})^T Q(x_{\kappa+1})(x - x_{\kappa+1}) \quad (8)$$

such that $Q(x_{k+1})$ is knowen as the Hessean_matrices. To get the minimum, differentiate the derivative by two sides of the Eq. (8) for $(x - x_{k+1})$, and equate to zero, we obtain:

$$g_{k+1} = g_k - Q(x_{k+1})(x - x_{k+1}) \quad (9)$$

Substituting x_k in to x in equation (8) and putting relation (9) into (8), we get:

$$1/2 s_k^T Q(x_{k+1}) s_k = (f_{k+1} - f_k - g_k^T s_k) \quad (10)$$

Now, by equation (10) and equation (3) we can see that:

$$s_k^T Q(x_{k+1}) s_k = (f_{k+1} - f_k) - 3/2 s_k^T g_k \quad (11)$$

We propose to use our $\tau_k I$ from an approximation to the Hessian, then above equation be:

$$\tau_k s_k^T s_k = (f_{k+1} - f_k) - 3/2 s_k^T g_k \quad (12)$$

From definition of y_k we have:

$$y_k^T s_k = g_{k+1}^T s_k - g_k^T s_k \tag{13}$$

By putting (12) in (11) imply:

$$\tau_k s_k^T s_k = \frac{3}{2} y_k^T s_k + (f_{k+1} - f_k) - \frac{3}{2} g_{k+1}^T s_k \tag{14}$$

will get the approximate optimal step-size:

$$\alpha_k = \tau_k^{-1} = \frac{3/2 y_k^T s_k + (f_{k+1} - f_k) - 3/2 g_{k+1}^T s_k}{s_k^T s_k} \tag{15}$$

Now, by using exact line search property in equation (15), we get:

$$\alpha_k = \tau_k^{-1} = \frac{3/2 y_k^T s_k + (f_{k+1} - f_k)}{s_k^T s_k} \tag{16}$$

As a result, we describe the algorithm for our proposed method and called Algorithms New1,2.

New Algorithm

St1. Select $x_0 \in R^n$ and compute $d_0 = -g_0$. Set $\kappa = 0$.

St 2. If converge case is satisy, then_stop.

St 3. Compute the step_length α_k as in Eqs. (15-16).

St 4. Update : $x_{k+1} = x_k - \alpha_k g_k$, then go to *St* 2.

3. Converge Proparty

This part focuses on analyzing the convergente property of the BFGS tacneque. Specifically, we make the assumption that f meets the requirements of being strong convex and the set $\Psi = \{x \in R^n : f(x) \leq f(x_0)\}$ was closed. The strong convex of f by Ψ is characterized by the presence of constants m and M, where:

$$mI \leq \nabla^2 f(x) \leq MI \tag{17}$$

to each $x \in \Psi$. A outcome for the strong convex of f on Ψ is must be bound f^* as :

$$f(x) - \frac{1}{2m} \|\nabla^2 f(x)\|_2^2 \leq f(x) \leq f(x) - \frac{1}{2M} \|\nabla^2 f(x)\|_2^2 \tag{18}$$

For more details see [3-4].

Theorm 1.

The new algorethm utilizing backtracking achieves linear convergence and

$$f(x_k) - f^* \leq \left(\prod_{i=0}^{k-1} c_i\right) f(x_0) - f^* \tag{19}$$

where $c_i = 1 - min\{m, ms^{p_k}\} < 1$and $p_k \geq 1$is an integer, ($p_k = 1,2,3,.,.,..$ as determined by the backtracking algorithm).

Proof :

By using (12), can rewrite :

$$\tau_k s_k^T s_k = (f_{k+1}\text{-}f_k)\text{-}3/2 s_k^T g_k \tag{20}$$

Now, in point $x_{\kappa+1} = x_\kappa - \alpha_k \nabla f(x_\kappa)$ and $d_k = -\nabla f(x_k)$ we have :

$$\tau_k \alpha_k^2 \nabla f(x_k)^T \nabla f(x_k) = (f_{k+1}\text{-}f_k)\text{-}3/2\alpha_k \nabla f(x_\kappa)^T \nabla f(x_\kappa) \tag{21}$$

Hence :

$$\alpha_k \|\nabla f(x_k)\|_2^2 = (f_{k+1}\text{-}f_k)\text{-}3/2\alpha_k \|\nabla f(x_k)\|_2^2$$
$$f_{k+1}\text{-}f_k = \text{-}1/2\alpha_k \|\nabla f(x_k)\|_2^2 \tag{22}$$

By using backtrack procedure terminats may $\alpha_k = 1$ or $\alpha_k = s^{p_k}$ such that p_k is any integer. So:

$$f(x_{k+1}) = f(x_k)\text{-} \min\{r, rs^{p_k}\} \|\nabla f(x_k)\|_2^2 \tag{23}$$

Having that for strong convexity functions $\|\nabla f(x_k)\|_2^2 \geq 2m(f(x_k) - f^*)$ it follows that :

$$f(x_{\kappa+1})\text{-}f^* \leq c_\kappa(f(x_\kappa)\text{-}f^*) \tag{24}$$

where $c_k = min\{rm, rms^{p_k}\}$. because $c_k < 1$ n then the sequence $f(x_k)$ is linear convergence, like a geometricly series, f^*.

4. Numerical Results

In this part of the study, numerical findings are presented and the New method's effectiveness is contrasted with traditional gradient descent (GD) techniques for addressing unconstrained problems commonly cited in previous research [1]. The evaluation is conducted utilizing standard optimization problems outlined in prior literature [2]. Other different test functions can be found in [3-4]. We stopp the algorithm when the stopping condition: $\|\nabla f(x)\| \leq 10^{-6}$or $\alpha_k|g_k^T d_k| \leq 10^{-20}|f_{k+1}|$ be satisfy. For the new with the gradient_descent (GD) algorithm, this parameter would chosed: $\alpha = 0.0001$ with $s = 0.8$.
In order to assess the numerical effectiveness of the (CG) methods being tested, details on the performance can be found in Figures 1 and 2. These charts are based on the criteria set by Dolan and Moré [8], which include the numbers of iteration and the numbers of functions evaluation. We can also see from Figure 1 and 2, The new techniques have effectively reached the desired outcomes and have demonstrated strong performance
It is clear that New methods are strong competitive with GS method and slightly better in some cases for all graphs in Figures 1 and 2 which include the iterations number and evaluation function.

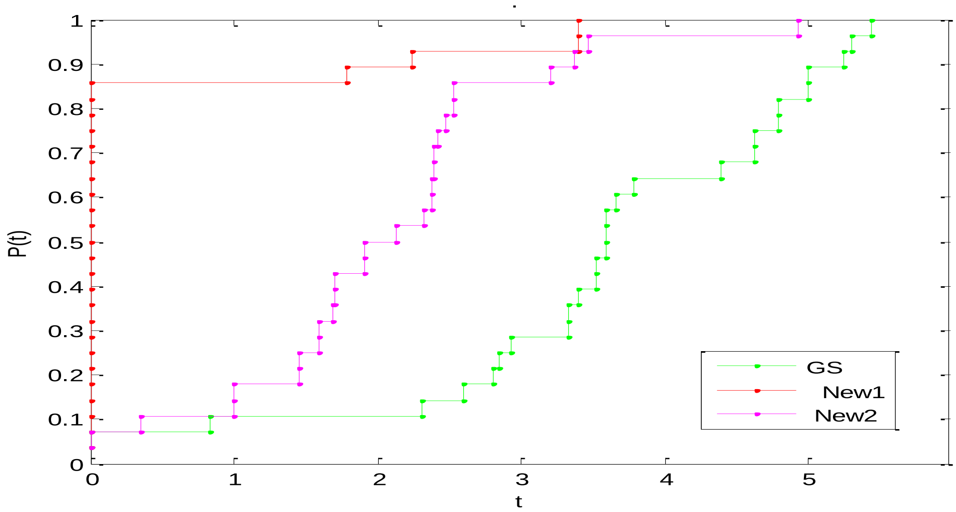

Figure 1: Profile of performance determined by the quantity of repetitions

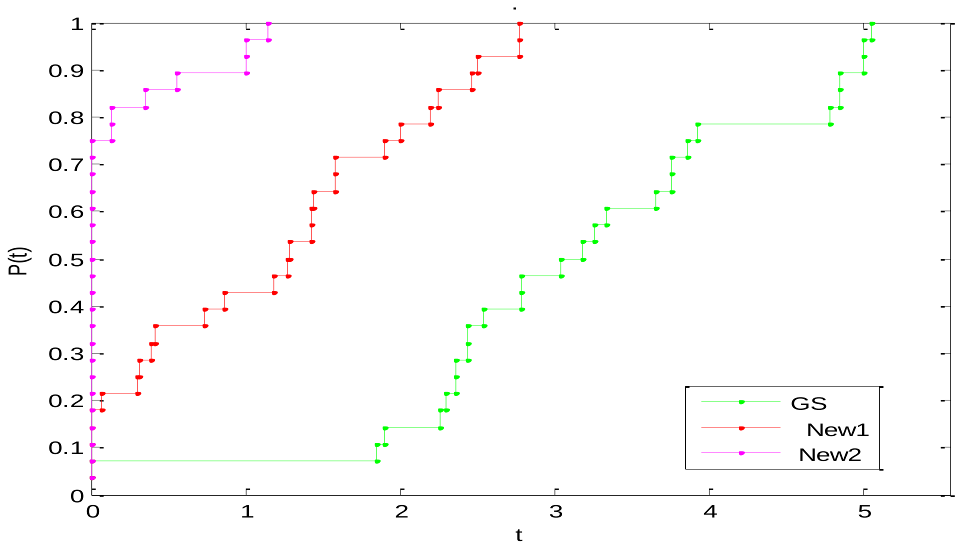

Figure 2: Assessment summary derived from changes in functionality

5. Conclusions

Our approach in this paper is based on utilizing a 2nd Taylor's series to determine the new step size. We have established the universal convergence characteristic of our method and shown through numerical trials that it is efficient. Additionally, we have developed alternative versions of our method using different higher-order tensor models.

Refrences

[1] Andrie N. (2008) ' An Unconstrained Optimization Test functions collection ' Advanced Modeling and optimization. 10, pp.147-161.

[2] Basim A. Hassan and Haneen A. Alashoor, (2023), On image restoration problems using new conjugate gradient methods, Indonesian Journal of Electrical Engineering and Computer Science Vol. 29, No. 3, March 2023, pp. 1438-1445.

[3] Basim A. Hassan and Hameed M. Sadiq, (2022), A new formula on the conjugate gradient method for removing impulse noise images, Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software (Bulletin SUSU MMCS), 2022, vol. 15, no. 4, pp. 123-130.

[4] Basim A. Hassan and Hameed M. Sadiq, (2022), Efficient New Conjugate Gradient Methods for Removing Impulse Noise Images, Europear Journal of Pure and Applied Mathematics, Vol. 15, No. 4, 2011-2021.

[5] Basim A. Hassan and Haneen A. Alashoor, (2022), A New Type Coefficient Conjugate on the Gradient Methods for Impulse Noise Removal in Images, Europear Journal of Pure and Applied Mathematics, Vol. 15, No. 4, 2022, 2043-2053.

[6] Basim A. Hassan and Ali Ahmed A. Abdullah, (2023), Improvement of conjugate gradient methods for removing impulse noise images, Indonesian Journal of Electrical Engineering and Computer Science, Vol. 29, No. 1, pp. 245-251.

[7] Biglari, F., Solimanpur, M., (2013), : Scaling on the spectral gradient method. J. Optim. Theory Appl. 158(2), pp.626–635.

[8] 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.

[9] Dolan, E.D., Moré, J.J. (2020), Benchmarking optimization software with performance profiles. Math. Program , 91(2), PP.201–213.

[10] Hager W., Zhang H.,(2006), A survey of nonlinear conjugate gradient methods, Pac. J. Optim. 2, pp. 35–58.

[11] H. N. Jabbar and Basim A. Hassan, “Two-versions of descent conjugate gradient methods for large-scale unconstrained optimization,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 22, no. 3, 2021,pp1643-1649.

[12] Rao, S.S. (2009). Optimization Theory and Applications. Wiley Eastern Limited.

[13] Nocedal J., Wright S.,(2006), Numerical Optimization, Second edition, Springer-Verlag, New York.

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

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

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