Membrane Technology

ISSN (online): 1873-4049

Utilization of Novel Memoryless Conjugate Gradient Algorithms for Nonlinear Unconstrained Optimization Challenges

Sabreen M. Abbas[1]; Abbas Y. Al-Bayati[2] & Maysoon M. Aziz[3]

[1] Department of Mathematics, College of Computer Sciences and Mathematics, University of Mosul, Iraq. e-mail:mssabreen6@gmail.com

[2] University of Telafer, Iraq.

e-mail:profabbasalbayati@yahoo.com

[3] University of Mosul, Iraq.

e-mail:Maysoon M.Aziz@gmail.com

Abstract.

This study introduces a novel family of memoryless nonlinear conjugate gradient algorithms that produce an appropriate search direction for gradient descent at each iteration. This condition is applicable irrespective of the exactness of the line search and the convexity of the goal function. We demonstrate that the offered approaches achieve global convergence for non-convex functions under specific conditions. Numerical findings illustrate the efficacy of these novel hybrid approaches when applied to specific test issues in comparison to the conventional Mawlana CG algorithm.

1. Introduction.

We are interested in the unconstrained optimization problem[1]

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

where $f: R^n \rightarrow R$ is a continuously differentiable function. We denote its gradient ∇f

by$g$. Quasi-Newton methods are known as effective numerical methods for solving problem (1), and they are iterative methods of the form[2]:

$$x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

Where x_k is the current iteration point and α_k is the step length being calculated

By performing a line search, d_k is the search direction specified by[3]:

$$d_{k+1} = \begin{cases} -g_{k+1} & if\ k = 0 \\ -g_{k+1} + \beta_k d_k & if\ k \geq 1 \end{cases} \tag{3}$$

where $\beta_k$is scalar. The differences highlight CG methods in their speed and performance by specifying the numerical parameter β_k, and the set of the effective versions of $\beta_k$acontained in source [4].

However, it is important to note that the behavior of the following approaches might vary significantly for general objective functions, even if they may appear equivalent in cases where f is a strictly convex quadratic function and k is determined using exact line searches (ELS). Wolfe conditions, strong Wolfe conditions, and strong Wolfe conditions are examples of inexact line searches (ILS) that are hoped for in the convergence analysis of conjugate gradient methods.

1. standard Wolfe line search:

$$\begin{cases} f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \\ d_k^T g(x_k + \alpha_k d_k) \geq \sigma d_k^T g_k \end{cases} \tag{4a}$$

2. strong Wolfe line search:

$$\begin{cases} f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \\ |d_k^T g(x_k + \alpha_k d_k)| \leq -\sigma d_k^T g_k \end{cases} \tag{4b}$$

2. Equations (2) can be represented by:

$$\begin{cases} f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k \\ \sigma d_k^T g_k \le d_k^T g(x_k + \alpha_k d_k) \le 0 \end{cases} \quad (5)$$

Where $0 < \delta < \frac{1}{2}$ and $\delta < \sigma < 1$ [5]

Optimization plays a pivotal role in various scientific, engineering, and economic applications. The pursuit of finding optimal solutions efficiently and accurately has led to the development of numerous mathematical methods and algorithms. Among these, the Quasi-Newton methods have garnered significant attention due to their balance of computational efficiency and convergence properties[6].
The traditional Quasi-Newton methods rely on approximating the Hessian matrix, which can be computationally expensive and memory-intensive for large-scale problems. To address these challenges, researchers have developed the Memoryless Quasi-Newton Method, which offers a promising alternative by reducing the memory requirements while maintaining robust performance in optimization tasks[7].

This paper aims to provide a comprehensive overview of the Memoryless Quasi-Newton Method, including its theoretical foundations, algorithmic structure, and practical applications. By eliminating the need for storing and updating the Hessian matrix or its approximations, this method significantly reduces the computational burden, making it particularly suitable for high-dimensional optimization problems.

The Memoryless Quasi-Newton Method represents a significant advancement in optimization techniques, balancing the need for accurate and efficient optimization with the practical constraints of memory and computational resources. By leveraging this method, practitioners can tackle large-scale optimization problems more effectively, opening the door to new possibilities and innovations in various fields[7], [8].

2 New Directions in Memoryless Methodologies

The memoryless variant addresses these issues by not explicitly storing or updating the Hessian matrix or its approximations.

- **The mathematical form**

A common memoryless Quasi-Newton method is the Memoryless BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm, where the search direction is updated as follows:

$$d_{k+1} = -\nabla f_{k+1} + \beta_k d_k \quad (6)$$

Here, β_k is a scalar that can be determined by various formulas.

Sometimes, conjugate gradient formulas produce the TTCG type.

$$d_{k+1} = -\nabla f_{k+1} + \beta_k d_k - \gamma y_k \quad (7)$$

And γ parameter where represent the third term parameter to produce TTCG.

Our approach to the memoryless direction of the three-term type of Hessian matrix commences with a mole of conjugate gradient formulations. Such as the parameter of conjugation for the researcher Mawlana at certain times and for revising the formulas of the researcher Mawlana at other times.

2.1 New Maulana Memoryless algorithm

The conjugate gradient and the preconditional conjugate gradient were discussed in the third and fourth chapters respectively. The inclusion of the two chapters was admirable, nevertheless, it exhibited some slowing down when dealing with high dimensions problems. Hence, our objective is to decrease memory consumption in the enhanced algorithms discussed in the preceding two chapters by implementing the notion of memoryless memory.

$$d_{k+1}^{NEWi} = -H_{k+1}^{BFGS} g_{k+1} + H_k^{BFGS} \beta_k^{M1} d_k$$

If we are discussing the preconditional tendency, Utilizing the concept of memory lessness, we deduct the Hessian approximation matrix H from the neutral convolution, thereby determining the direction using the following formula:

$$d_{k+1}^{NEWi} = -H_{k+1}^{BFGS} g_{k+1} + H_k^{BFGS} \beta_k^{M1} d_k \quad (8)$$

$$d_{k+1}^{NEWi} = -H_{k+1}^{BFGS} g_{k+1} + H_k^{BFGS} \beta_k^{M2} d_k \quad (9)$$

$$H_{k+1}^{BFGS} = \left(I - \frac{y_k s_k^T + s_k y_k^T}{y_k^T s_k} \right) + \left(\left(1 + \frac{y_k^T y_k}{y_k^T s_k} \right) \frac{s_k s_k^T}{y_k^T s_k} \right)$$

$$x_{k+1} = x_k + \alpha_k d_k$$

α_k here is close to ensure that

$s_k^T d_k > 0$, Hence H_{k+1}^{BFGS} is +ive definite

$d_{k+1}^T g_{k+1} < 0$

After placing I on the matrix H, Where the parameter of conjugacy β_k^{M1} , β_k^{M2} given by

$$\beta_k^{M1} = \frac{{g_{k+1}}^T \left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_{k-1}\|} g_k - g_k \right)}{{g_k}^T (g_{k+1} - d_k)} \quad (10)$$

$$\beta_k^{M2} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_{k+1} - g_k\|} |{g_{k+1}}^T g_k| - |{g_{k+1}}^T g_k|}{(1-\mu)\|d_k\|^2 + \mu \|g_k\|^2} \quad (11)$$

Where $\mu = 0.6$

2.2 Our memoryless algorithms Outline (M1)

1- For $X_0 \in R^n$, $0 < \in << 1, 0 < \delta < \frac{1}{2}$, *and* $\delta < \sigma < 1$,
2- set $d_0 = -g_0$, $k = 0$.
3- If $\|g_k\| < \in$, then stop, otherwise go to the next step.
4- Compute step size α_k by Wolfe line search (4(a,b))
5- Let $x_{k+1} = x_k + \alpha_k d_k$, if $\|g_{k+1}\| < \in$, then stop.
6- Calculate the new search directions by

$$d_{k+1}^{NEWi} = -H_{k+1}^{BFGS} g_{k+1} + H_k^{BFGS} \beta_k^{M1} d_k$$

$$\beta_k^{M1} = \frac{{g_{k+1}}^T \left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_{k-1}\|} g_k - g_k \right)}{{g_k}^T (g_{k+1} - d_k)}$$

7- Set $k = k + 1$, and go to step2.

Property (Gilbert and J. Nocedal, 1992)

Suppose that the general conjugate gradient method is used and that $0 < \delta \le \|g_k\| \le \bar{\delta}$ is achieved in it, then this method has property if the constants $b > 1$ and $p > 1$ are found, for example, for every k,

$$|\beta_k| \leq b \tag{12}$$

$$\|s_k\| \leq p \Rightarrow |\beta_k| \leq \frac{1}{2b} \tag{13}$$

2.1.3 Theorem

Suppose that Assumption (3.3.1) (3.3.2) holds. Any CG method of the form (2) and (3) with d_k is a descent search direction and α_k satisfies (SWP) in (4b). If

$$\sum_{k\geq 0} \frac{1}{\|d_k\|^2} = \infty \Rightarrow \lim_{k\to\infty} \inf\|g_k\| = 0 \tag{14}$$

(Dai and Liao, 2001)

2.3 The Descent Property of a CG New Method ($M1$)

The descent property for our proposed new conjugate gradient scheme must be demonstrated below, referred to as $M1$. In the next, we argue the sufficient descent.

2.3.1 *Theorem (1)*

The search direction d_{k+1} and β_k^{M1} given in equation

$$\mathrm{d}_{\mathrm{k+1}}^{\mathrm{NEWi}} = -H_{k+1}^{BFGS}\mathrm{g}_{\mathrm{k+1}} + H_k^{BFGS}\beta_{\mathrm{k}}^{\mathrm{M1}}\mathrm{d}_{\mathrm{k}} \tag{15}$$

is a descent direction where:

$$H_{k+1}^{BFGS} = \left(\mathrm{I} - \frac{y_k s_k^T + s_k y_k^T}{y_k^T s_k}\right) + \left(\left(1 + \frac{y_k^T y_k}{y_k^T s_k}\right)\frac{s_k s_k^T}{y_k^T s_k}\right)$$

From Lipschitz condition we have

$$\|y_k\| = l\|s_k\|$$

$$H_{k+1}^{BFGS} = \left(\mathrm{I} - \frac{l\|s_k\|^2 + l\|s_k\|^2}{l\|s_k\|^2}\right) + \left(\left(1 + \frac{l^2\|s_k\|^2}{l\|s_k\|^2}\right)\frac{\|s_k\|^2}{l\|s_k\|^2}\right)$$

$$H_{k+1}^{BFGS} = \left(\mathrm{I} - \frac{2l\|s_k\|^2}{l\|s_k\|^2}\right) + \left(\frac{\|s_k\|^2}{l\|s_k\|^2} + 1\right)$$

$$H_{k+1}^{BFGS} = \left(\mathrm{I} - 1 + \frac{1}{l}\right) = R_k \tag{16}$$

$$H_k^{BFGS} = \left(\mathrm{I} - 1 + \frac{1}{l}\right) = R_{k-1} \tag{17}$$

where I is Diagonal matrix

and

$$\beta_k^{M1} = \frac{g_{k+1}{}^T\left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_{k-1}\|}g_k - g_k\right)}{g_k{}^T(g_{k+1} - d_k)} \tag{18}$$

Proof: We begin by Multiplying (15) by $\mathrm{g}_{\mathrm{k+1}}$ and substituting H_{k+1}^{BFGS}, H_k^{BFGS} by (16), (17) in (15)

$$d_{k+1}^T g_{k+1} = -R_k \|g_{k+1}\|^2 + R_{k-1} \beta_k^{M1} d_k^T g_{k+1} \quad (19)$$

By using (IEL) property, we get

$$d_k^T g_{k+1} = d_k^T g_{k+1} - d_k^T g_k + d_k^T g_k$$

$$= d_k^T (g_{k+1} - g_k) + d_k^T g_k = d_k^T y_k + d_k^T g_k < d_k^T y_k \quad (20)$$

Subsuming(20) in (19)

$$d_{k+1}^T g_{k+1} = -R_k \|g_{k+1}\|^2 + R_{k-1} \beta_k^{M1} d_k^T y_k$$

$$d_{k+1}^T g_{k+1} = -R_k \|g_{k+1}\|^2 + \frac{g_{k+1}{}^T \left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k - g_k \right)}{g_k{}^T (g_{k+1} - d_k)} R_{k-1} d_k^T y_k$$

$$d_{k+1}^T g_{k+1} = -R_k \|g_{k+1}\|^2 + \left(\frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} g_{k+1}^T g_k - g_{k+1}^T g_k}{g_k{}^T g_{k+1} - g_k{}^T d_k} \right) R_{k-1} d_k^T y_k$$

by powell condition

$$g_{k+1}^T g_k = 0.2 \|g_{k+1}\|^2$$

$$d_{k+1}^T g_{k+1} = -R_k \|g_{k+1}\|^2 + \left(\frac{\|g_{k+1}\|^2 + \frac{0.2\|g_{k+1}\|^3}{\|g_k\|} + 0.2\|g_{k+1}\|^2}{g_k{}^T g_{k+1} - g_k{}^T d_k} \right) R_{k-1} d_k^T y_k$$

By using descent condition

$$g_k{}^T d_k \le 0$$

$$d_{k+1}^T g_{k+1} \le -R_k \|g_{k+1}\|^2 + \left(\frac{\frac{\|g_k\| \|g_{k+1}\|^2 + 0.2\|g_{k+1}\|^3 + 0.2\|g_k\| \|g_{k+1}\|^2}{\|g_k\|}}{g_k{}^T g_{k+1}} \right) R_{k-1} d_k^T y_k$$

$$d_{k+1}^T g_{k+1} \le -R_k \|g_{k+1}\|^2 + \left(\frac{\|g_k\| + 0.2\|g_{k+1}\| + 0.2\|g_k\|}{\|g_k\| g_k{}^T g_{k+1}} \right) \|g_{k+1}\|^2 R_{k-1} d_k^T y_k$$

$$d_{k+1}^T g_{k+1} \le -R_k \|g_{k+1}\|^2 + \left(\frac{1.2\|g_k\| + 0.2\|g_{k+1}\|}{\|g_k\| g_k{}^T g_{k+1}} \right) \|g_{k+1}\|^2 R_{k-1} d_k^T y_k \quad (21)$$

$$d_{k+1}^T g_{k+1} \le -R_k \|g_{k+1}\|^2 + \left(\frac{1.2\|g_k\| + 0.2\|g_{k+1}\|}{\|g_k\| g_k{}^T g_{k+1}} \right) \|g_{k+1}\|^2 R_{k-1} d_k^T y_k$$

$$d_{k+1}^T g_{k+1} \le -\left(R_k + \left(\frac{1.2\|g_k\| + 0.2\|g_{k+1}\|}{\|g_k\| g_k{}^T g_{k+1}} \right) R_{k-1} d_k^T y_k \right) \|g_{k+1}\|^2$$

$$d_{k+1}^T g_{k+1} \le -c \|g_{k+1}\|^2$$

Where

$$c = R_k + \left(\frac{1.2\|g_k\| + 0.2\|g_{k+1}\|}{\|g_k\| g_k{}^T g_{k+1}} \right) R_{k-1} d_k^T y_k \quad , \quad c > 0$$

3 Global convergence study

We will show that the Memoryless CG technique with β_k^{M1} globally converges. For the suggested new algorithm's convergence, we need to make some assumptions.

3.1 Assumption (1)

a. Suppose f in the level set below, is bound

$K = \{x \in R^n : \varphi(x) \le \varphi(x_0)\}$; in some initial point.

b. φ is Its gradient is Lipschitz constant and it is continuously differentiable, there exists $L > 0$ in the sense that [9]:

$$\| g(x) - g(y) \| \le L\|x - y\| \forall x, y \in N \quad (22)$$

Under assumption(1), on the other hand, it is obvious that a positive constant M exists

$$\|x\| \le M, \forall x \in K \quad (23)$$

$$\| \nabla f(x) \| \le \bar{\gamma}, \forall x \in K \quad (24)$$

3.1.1 *Lemma*

Assume Assumption (1) and Equation (18) are correct. Take any conjugate gradient approach between (2) and (3), where d_k is the descent direction and α_k is the S.W.L.S. If

$$\sum_{k>1} \frac{1}{\| d_{k+1} \|^2} = \infty$$

then we have

$$\lim_{k\to\infty} inf \| g_k \| = 0$$

More details can be found in [10] [11].

3.1.2 *Theorem (2)*

Assume that Assumption (1) and Equation (2) are true, as well as the descent condition. Consider the following conjugate gradient scheme:

$$\mathrm{d}_{k+1}^{\mathrm{NEWi}} = -H_{k+1}^{BFGS} \mathrm{g}_{k+1} + H_k^{BFGS} \beta_k^{M1} \mathrm{d}_k$$

Where α_k is calculated based on the strong Wolfe line search condition (SWLS); for more details see[12][13]. If the objective function is uniformly on set S, then

$$\lim_{n\to\infty} (inf \| g_k \|) = 0 .$$

Proof

$$\mathrm{d}_{k+1}^{\mathrm{NEWi}} = -H_{k+1}^{BFGS} \mathrm{g}_{k+1} + H_k^{BFGS} \beta_k^{M1} \mathrm{d}_k$$

$$\|d_{k+1}\| = R_k \|g_{k+1}\| + R_{k-1} |\beta_k^{M1}| \|d_k\| \quad (25)$$

We take each part separately and deduce its value:

$$\beta_k^{M1} = \frac{{g_{k+1}}^T \left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_{k-1}\|} g_k - g_k \right)}{{g_k}^T (g_{k+1} - d_k)}$$

$$\beta_k^{M1} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{\|g_k\|} g_k - g_{k+1}^T g_k}{{g_k}^T g_{k+1} - {g_k}^T d_k}$$

$$|\beta_k^{M1}| = \left| \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{\|g_k\|} g_k - g_{k+1}^T g_k}{g_k{}^T g_{k+1} - g_k{}^T d_k} \right|$$

$$|\beta_k^{M1}| \le \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{\|g_k\|} \|g_k\| - \|g_{k+1}\|\|g_k\|}{\|g_k\|\|g_{k+1}\| - \|g_k\|\|d_k\|}$$

$$|\beta_k^{M1}| \le \frac{-\|g_{k+1}\|\|g_k\|}{\|g_k\|\|g_{k+1}\| - \|g_k\|\|d_k\|}$$

$$|\beta_k^{M1}| \le \frac{-\bar{\gamma}\bar{\delta}}{\bar{\delta}\,\bar{\gamma} - \bar{\delta}\,\|d_k\|}$$

$$|\beta_k^{M1}| \le \frac{-\bar{\gamma}\bar{\delta}}{-\bar{\delta}\,(\|d_k\| - \bar{\gamma})}$$

$$|\beta_k^{M1}| \le \frac{\bar{\gamma}}{\|d_k\| - \bar{\gamma}} = E_1 \tag{26}$$

Substituting the (26) in(25)

$\|d_{k+1}\| = R_k\bar{\gamma} + R_{k-1}E_1\|d_k\|$

$\|d_{k+1}\| = R_k\bar{\gamma} + R_{k-1}E\|d_k\|$

Then we get

$$0 < \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty$$

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} \le \sum_{k=0}^{\infty} \frac{1}{c^2} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty$$

then

$$\lim_{k\to\infty} \inf \|g_k\| = 0.$$

3.1.3Our memoryless algorithms Outline (M2)

1- For $X_0 \in R^n$, $0 < \in << 1, 0 < \delta < \frac{1}{2}$, *and* $\delta < \sigma < 1$,
2- set $d_0 = -g_0$, $k = 0$.
3- If $\|g_k\| < \in$, then stop, otherwise go to the next step.
4- Compute step size α_k by Wolfe line search (4(a,b)).
5- Let $x_{k+1} = x_k + \alpha_k d_k$, if $\|g_{k+1}\| < \in$, then stop.
6- Calculate the new search directions by

$$\mathrm{d}_{k+1}^{\mathrm{NEWi}} = -H_{k+1}^{BFGS}\mathrm{g}_{k+1} + H_k^{BFGS}\beta_k^{M2}\mathrm{d}_k$$

$$\beta_k^{M2} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_{k+1} - g_k\|} |g_{k+1}{}^T g_k| - |g_{k+1}{}^T g_k|}{(1-\mu)\|d_k\|^2 + \mu\|g_k\|^2}$$

7- Set $k = k + 1$, and go to step2.

4 The Descent Property of a CG New Method(M2)

The descent property for our proposed new conjugate gradient scheme must be demonstrated below, referred to as $M2$. In the next, we argue the sufficient descent.

Proof:

Starting by the direction of Memoryless (4)

$$d_{k+1}^{NEWi} = -H_{k+1}^{BFGS} g_{k+1} + H_k^{BFGS} \beta_k^{M2} d_k \quad (27)$$

$$\beta_k^{M2} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_{k+1} - g_k\|} |g_{k+1}{}^T g_k| - |g_{k+1}{}^T g_k|}{0.4\|d_k\|^2 + 0.6\|g_k\|^2} \quad (28)$$

where μ =0.6

Multiply (27) by g_{k+1} and substituting $H_{k+1}^{BFGS}, H_k^{BFGS}$ by (16), (17) in (27)

$$d_{k+1}^T g_{k+1} = -R_k \|g_{k+1}\|^2 + R_{k-1} \beta_k^{M2} d_k^T g_{k+1} \quad (29)$$

By using (IEL)

$$d_k^T g_{k+1} = d_k^T g_{k+1} - d_k^T g_k + d_k^T g_k$$

$$= d_k^T (g_{k+1} - g_k) + d_k^T g_k = d_k^T y_k + d_k^T g_k < d_k^T y_k \quad (30)$$

Substituting (30) in (29)

$$d_{k+1}^T g_{k+1} = -R_k \|g_{k+1}\|^2 + R_{k-1} \beta_k^{M1} d_k^T y_k$$

$$d_{k+1}^T g_{k+1} = -R_k \|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_{k+1} - g_k\|} |g_{k+1}{}^T g_k| - |g_{k+1}{}^T g_k|}{0.4\|d_k\|^2 + 0.6\|g_k\|^2} R_{k-1} d_k^T y_k$$

$$d_{k+1}^T g_{k+1} = -R_k \|g_{k+1}\|^2 + \left(\frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_{k+1} - g_k\|} |g_{k+1}{}^T g_k| - |g_{k+1}{}^T g_k|}{0.4\|d_k\|^2 + 0.6\|g_k\|^2} \right) R_{k-1} d_k^T y_k$$

$$d_{k+1}^T g_{k+1} = -R_k \|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|y_k\|} |g_{k+1}{}^T g_k| - |g_{k+1}{}^T g_k|}{0.4\|g_k\|^2 + 0.6\|g_k\|^2} R_{k-1} d_k^T y_k$$

$$d_{k+1}^T g_{k+1} \le -R_k \|g_{k+1}\|^2 + \frac{\frac{\|y_k\|\|g_{k+1}\|^2 - \|g_{k+1}\|\|g_{k+1}{}^T g_k| - \|y_k\|\|g_{k+1}{}^T g_k|}{\|y_k\|}}{\|g_k\|^2} R_{k-1} d_k^T y_k$$

By powell condition we get

$$d_{k+1}^T g_{k+1} \le -R_k \|g_{k+1}\|^2 + \frac{\frac{l\|s_k\|\|g_{k+1}\|^2 + 0.2\|g_{k+1}\|^3 + 0.2l\|s_k\|\|g_{k+1}\|^2}{l\|s_k\|}}{\|g_k\|^2} R_{k-1} d_k^T y_k$$

$$d_{k+1}^T g_{k+1} \le -R_k \|g_{k+1}\|^2 + \left(\frac{l\|s_k\| + 0.2\|g_{k+1}\| + 0.2l\|s_k\|}{l\|s_k\|\|g_k\|^2} \right) \|g_{k+1}\|^2 R_{k-1} d_k^T y_k \quad (31)$$

$$d_{k+1}^T g_{k+1} \le -R_k \|g_{k+1}\|^2 + \left(\frac{1.2l\|s_k\| + 0.2\|g_{k+1}\|}{l\|s_k\|\|g_k\|^2} \right) \|g_{k+1}\|^2 R_{k-1} d_k^T y_k$$

$$d_{k+1}^T g_{k+1} \le -\left(R_k - \left(\frac{1.2l\|s_k\| + 0.2\|g_{k+1}\|}{l\|s_k\|\|g_k\|^2} \right) R_{k-1} d_k^T y_k \right) \|g_{k+1}\|^2$$

$$d_{k+1}^T g_{k+1} \le -c\|g_{k+1}\|^2$$

where

$c > 0$.

5 Theorem (Global convergence)

Assume that Assumption (1) and Equation (1.33) are true, as well as the descent condition. Consider the following conjugate gradient scheme:

$$d_{k+1}^{NEWi} = -H_{k+1}^{BFGS} g_{k+1} + H_k^{BFGS} \beta_k^{M2} d_k$$

Where α_k is calculated based on the strong Wolfe line search condition (SWLS). If the objective function is uniformly on set S, then

$$\lim_{n\to\infty}(\inf \| g_k \|) = 0 .$$

Proof

$$d_{k+1}^{NEWi} = -H_{k+1}^{BFGS} g_{k+1} + H_k^{BFGS} \beta_k^{M2} d_k$$
$$\left|d_{k+1}^{NEWi} = -H_{k+1}^{BFGS} g_{k+1} + H_k^{BFGS} \beta_k^{M2} d_k\right|$$

$$\|d_{k+1}\| = R_k\|g_{k+1}\| + R_{k-1}|\beta_k^{M2}|\|d_k\| \qquad (32)$$

We take each part separately and deduce its value:

$$\beta_k^{M2} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_{k+1} - g_k\|}|g_{k+1}{}^T g_k| - |g_{k+1}{}^T g_k|}{(1-\mu)\|d_k\|^2 + \mu\|g_k\|^2}$$

where μ =0.6

$$\left|\beta_k^{M2} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_{k+1} - g_k\|}|g_{k+1}{}^T g_k| - |g_{k+1}{}^T g_k|}{0.4\|d_k\|^2 + 0.6\|g_k\|^2}\right|$$

$$|\beta_k^{M2}| \le \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_{k+1} - g_k\|}\|g_{k+1}\|\|g_k\| - \|g_{k+1}\|\|g_k\|}{0.4\|d_k\|^2 + 0.6\|g_k\|^2}$$

$$|\beta_k^{M2}| \le \frac{\frac{\|y_k\|\|g_{k+1}\|^2 - \|g_{k+1}\|^2\|g_k\| - \|y_k\|\|g_{k+1}\|\|g_k\|}{\|y_k\|}}{\|g_k\|^2}$$

$$|\beta_k^{M2}| \le \frac{\|y_k\|\|g_{k+1}\|^2 - \|g_{k+1}\|^2\|g_k\| - \|y_k\|\|g_{k+1}\|\|g_k\|}{\|y_k\|\|g_k\|^2}$$

$$|\beta_k^{M2}| \le \frac{l\|s_k\|\|g_{k+1}\|^2 - \|g_{k+1}\|^2\|g_k\| - l\|s_k\|\|g_{k+1}\|\|g_k\|}{l\|s_k\|\|g_k\|^2}$$

Let $\|s_k\| = D$

$$|\beta_k^{M2}| \le \frac{lD\bar{\gamma}^2 - \bar{\gamma}^2\bar{\delta} - lD\bar{\gamma}\bar{\delta}}{lD\bar{\delta}^2}$$

$$|\beta_k^{M2}| \le \frac{lD\bar{\gamma}(\bar{\gamma} - \bar{\delta}) - \bar{\gamma}^2\bar{\delta}}{lD\bar{\delta}^2} = E_2 \qquad (33)$$

Substituting the (33) in(32)

$$\|d_{k+1}\| = R_k\bar{\gamma} + R_{k-1}E_2\|d_k\|$$

$$\|d_{k+1}\| = R_k\bar{\gamma} + R_{k-1}E_2\|d_k\|$$

Then we get

$$0 < \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty$$

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} \leq \sum_{k=0}^{\infty} \frac{1}{c^2} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty$$

then

$$\lim_{k\to\infty} \inf \|g_k\| = 0.$$

6 Numerical examples

In this section, we use the test problem considered in (Andrei 2008)

We test and compare standard algorithm with the new memoryless algorithm Maulana1(M1) and algorithmMaulana2 (M2), in following manners:

- Maulana1 & ML-Maulana1with BFGS and Maulana2 & ML-Maulana2with BFGS, and the details in **table (1)**

For each of large-scale optimization problems, we have chosen unique test functions. These modified Visual FORTRAN CG algorithms are enforced using the Wolfe-Powell line search technique. The Intel(R) Core(TM) i7-6700HQ processor in a single laptop was used for all the computational experiments, Core TM, i7-3612QM (2.10GHZ) central processing unit, 6GB of random access memory, and Windows 10.

varying dimensions, arriving at 1,000 variables, representing a large scale for the descent approach. This expresses a total of Nxn {n=dimension and N is the problem} test problem so that it will arrive at N=60 and n=1000 for our experience. The number of iterations overextends its limit, which is set to be 1,000. In our implementation, the numerical tests were performed on an HP with an XP operating system and using Fortran to run the programming for both methods and switch approaches. The discussion of the whole comparison to estimate the variation between practices is given by Dolan and More using Matlab codes for this purpose, where comparisons are made using curves to demonstrate who is superior to whom based on the comparison's foundations and criteria. We listed 40 large scale unconstrained optimization test functions in generalized or extended. All methods implemented with Wolfe line search conditions with $\rho = 0.0001$ and same stopping criterion.

$\|g_k\|_2^2 \leq 10^{-6}$.

The comparison of the algorithms is based on the number of iterations (NOI), the number of function evaluations(NOFG) and (TIME) using Donald and More performance profiles.

Table (6.1). Comparison results by (Maulana1 & ML-Maulana1with BFGS) and (Maulana2 & ML-Maulana2 with BFGS).

Test Problems	**Maulana1** NOI/NOFG/ TIME	**ML-Maulana1with BFGS** NOI/NOFG/ TIME	**Maulana2** NOI/NOFG/ TIME	**ML-Maulana2with BFGS** NOI/NOFG/ TIME
1- Trigonometric	95/192/0.21	107/206/0.03	104/221/0.05	102/197/0.04

ISSN (online): 1873-4049

2- Extended Rosenbrock (CUTE)	1760/2674/0.36	170/387/0.00	932/1621/0.05	177/416/0.01
3- Extended White & Holst	4004/4809/0.35	181/462/0.01	2693/3512/0.09	187/452/0.00
4- Extended Beale	357/728/0.01	48/101/0.00	187/471/0.02	48/101/0.00
5- Raydan 2	48/133/0.01	12/36/0.02	48/133/0.01	12/36/0.00
6- Extended Tridiagonal 1	606/960/0.03	38/85/0.00	195/500/0.02	38/85/0.00
7- Extended Three Expo Terms	78/172/0.02	130/2770/0.43	88/172/0.03	105/1930/0.34
8- Generalized Tridiagonal 2	2121/2774/0.17	254/432/0.01	2101/2488/0.14	283/487/0.01
9- Diagonal 4	247/521/0.02	16/40/0.02	145/397/0.01	16/40/0.02
10- Diagonal 5	32/112/0.02	12/36/0.00	32/112/0.01	12/36/0.00
11- Extended Himmelblau	82/206/0.00	38/80/0.01	85/207/0.01	38/80/0.00
12- Extended PSC1	78/183/0.02	25/62/0.02	79/184/0.03	24/60/0.02
13- Extended BD1	112/248/0.01	252/402/0.03	104/235/0.01	128/267/0.01
14- Extended Hiebert	4004/4340/0.16	325/719/0.00	3942/5180/0.15	465/1263/0.03
15- Extended EP1	26/125/0.00	2/16/0.00	26/125/0.00	2/16/0.00

16- Extended Tridiagonal 2	203/386/0.00	209/754/0.03	189/375/0.00	183/610/0.04
17- ARROWHEAD (CUTE)	382/758/0.08	39/242/0.00	194/538/0.01	36/198/0.00
18- NONDIA (CUTE)	3533/4053/0.19	41/90/0.00	2602/2848/0.13	43/95/0.00
19- DQDRTIC (CUTE)	453/854/0.02	381/723/0.02	256/580/0.01	390/732/0.03
20- DIXMAANA (CUTE)	57/144/0.00	25/58/0.00	57/144/0.01	27/65/0.00
21- DIXMAANB (CUTE)	56/141/0.00	39/75/0.00	56/141/0.01	38/73/0.02
22- DIXMAANC (CUTE)	61/152/0.01	51/99/0.00	61/152/0.02	51/99/0.00
23- Broyden Tridiagonal	370/683/0.03	159/276/0.01	408/817/0.03	171/290/0.01
24- Tridiagonal Perturbed Quadratic	1332/2345/0.10	53/153/0.02	1223/2524/0.06	53/153/0.02
25- LIARWHD (CUTE)	1726/2728/0.10	77/178/0.00	1300/1931/0.08	70/164/0.00
26- DIAGONAL 6	48/133/0.01	12/36/0.02	48/133/0.01	12/36/0.00
27- DENSCHNA (CUTE)	108/202/0.02	46/89/0.00	90/189/0.01	46/89/0.00
28- DENSCHNC (CUTE)	207/434/0.03	49/100/0.01	164/387/0.05	44/95/0.00

29- DENSCHNB (CUTE)	57/157/0.00	24/60/0.00	57/157/0.01	20/52/0.02
30- DENSCHNF (CUTE)	91/212/0.01	75/141/0.00	90/206/0.00	81/161/0.00
31- Extended Block-Diagonal BD2	230/453/0.03	44/90/0.01	164/359/0.02	36/84/0.00
32- Generalized quartic GQ1	57/158/0.01	25/70/0.00	56/156/0.00	25/70/0.00
33- DIAGONAL 7	40/127/0.01	11/42/0.02	40/127/0.02	11/42/0.00
34- Full Hessian	45/165/0.02	8/28/0.00	45/165/0.02	8/28/0.00
35- SINCOS	78/183/0.02	25/62/0.01	79/184/0.02	24/60/0.01
36- Generalized quartic GQ2	229/442/0.01	177/297/0.00	186/379/0.01	189/316/0.02
37- ARGLINB (CUTE)	4/12/0.00	0/12/0.00	4/12/0.00	0/12/0.00
38- FLETCHCR (CUTE)	179/355/0.01	118/226/0.00	157/312/0.02	119/225/0.01
39- HIMMELBG (CUTE)	32/40/0.01	32/44/0.02	32/40/0.00	32/44/0.00
40- HIMMELBH (CUTE)	64/164/0.00	24/56/0.00	72/168/0.01	24/56/0.00
Total	**23.292/33.658/2.11**	**3.354/9.835/0.75**	**18.391/28.572/1.19**	**3.370/9.315/0.66**

ISSN (online): 1873-4049

Total Work= NOI+NOFG+TIME	**59.06**	**13.939**	**48.153**	**13.345**

Table(6.2) Percentage performance of (ML-Maulana1with BFGS) against (Maulana1) algorithms

TOOLS	Maulana1	ML-Maulana1with BFGS
NOI	100%	14.4%
NOFG	100%	29.2%
TIME	100%	35.6%

Table(6.3) No. of best NOFG test problems

Tools	No. of best NOFG (Maulana1)	No. of best NOFG (ML-Maulana1with BFGS)	No. of equal NOFG In both
NOI	3	35	2
NOFG	5	34	1
TIME	7	23	10

Table(6.4) Percentage performance of (ML-Maulana2with BFGS) against (Maulana2) algorithms

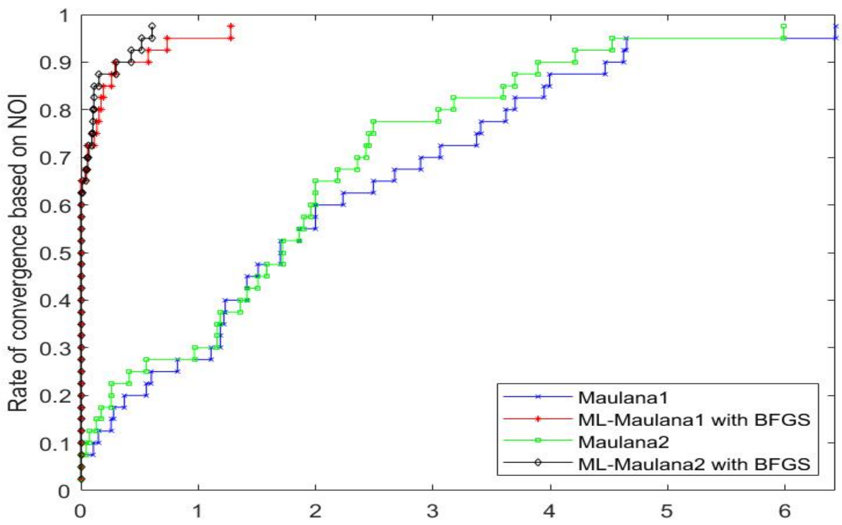
TOOLS	Maulana2	ML-Maulana2with BFGS
NOI	100%	18.3%
NOFG	100%	32.6%
TIME	100%	55.5%

Table(6.5) No. of best NOFG test problems

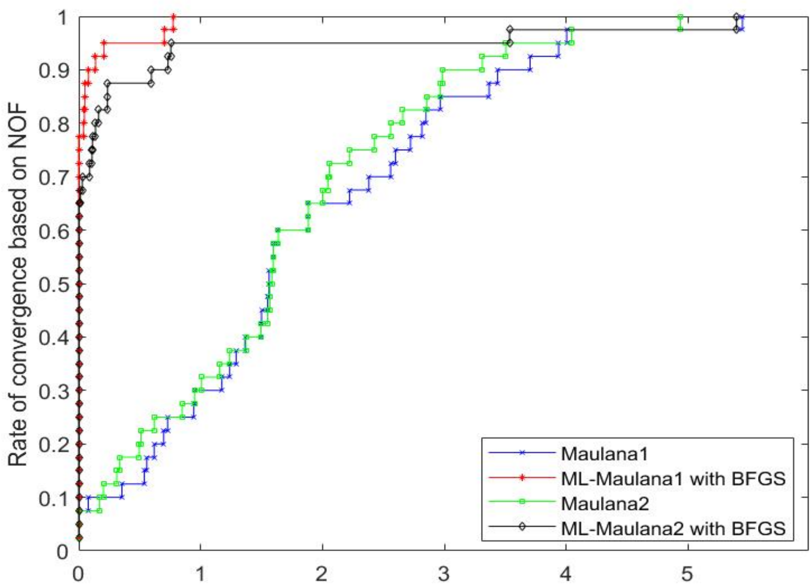
Tools	No. of best NOFG (Maulana2)	No. of best NOFG (ML-Maulana2with BFGS)	No. of equal NOFG In both
NOI	4	35	1
NOFG	5	34	0

TIME	**7**	**27**	**6**

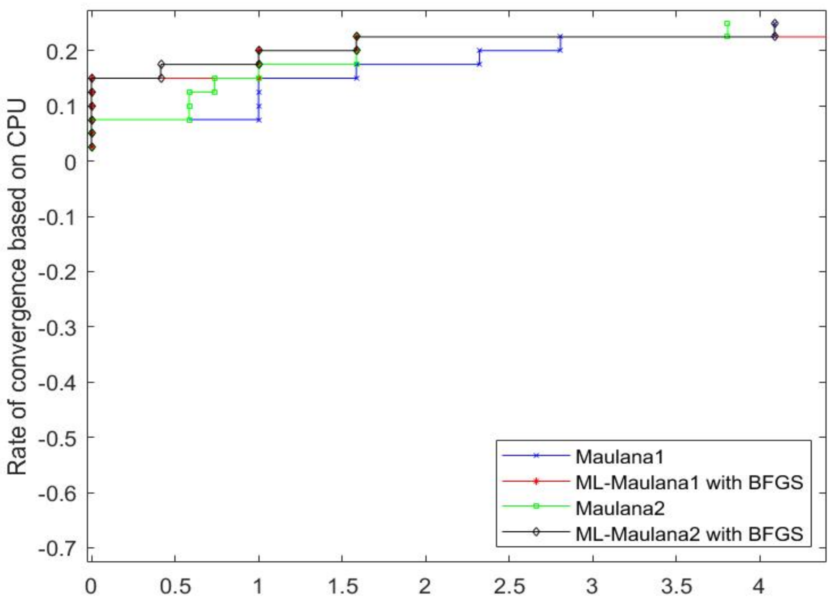
Tables(6.1-6.5) show that compared to the baseline Maulana1 CGmethod, the (ML-Maulana1with BFGS) algorithm improves upon it by a factor of (**14.4%,29.2%**, and **35.6%)** in terms of NOI, NOFG, and TIME. Comparing the (ML-Maulana2with BFGS) algorithm improves upon it by a factor of **(18.3%, 32.6%, 55.5%)** in terms of NOI, NOFG, and CPU, as shown in Table(6.3-6.5), reveals that the (ML-Maulana1and ML-Maulana2) algorithm achieves the strongest results in NOI, NOFG, and TIME under the accelerated Wolfe-Powell line search, demonstrating that the (ML-Maulana1and ML-Maulana2) algorithm is significantly more effective than the (Maulana1 and Maulana2) CG-algorithm. The (ML-Maulana1and ML-Maulana2) algorithm achieves its best performance by making full use of all available resources (including NOI, NOFG, and TIME.



Figure(6.1): Performance profile of (Maulana1 against ML-Maulana1with BFGS) relative to the NOI



Figure(6.2): Performance profile of (Maulana1 against ML-Maulana1with BFGS) relative to the NOF



Figure(6.3): Performance profile of (Maulana1 against ML-Maulana1with BFGS) relative to the TIME

Following the tables ((6.1)-(6.5)) and the graphs described in the performance file ((6.1)-(6.3)), the comparative findings that we carried out demonstrated a clear superiority in the functions that were chosen for the algorithms.

The nested and memory-enhanced algorithms, which are known as memoryless (Maulana1-BFGS) and memory-enhanced (Maulana2-BFGS), perform better than the classical algorithms (Maulana1 and Maulana2) when it comes to conjugate gradient methods.

Conclusion

We presented a unique approach that does not require memory and was developed particularly for issues that are not limited. The numerical results proved that the suggested algorithms provided higher efficiency in comparison to the algorithms ML-Maulana1 with BFGS and ML-Maulana2 with BFGS. The algorithms ML-Maulana1 with BFGS and ML-Maulana2 with BFGS displayed the best performance. When compared to the conventional Mawlana CG algorithm, the numerical results reveal that these novel hybrid approaches are just as successful when applied to a selection of testing problems.

References

[1] N. H. Al-Bayati, A.Y. and Al-Assady, “Conjugate Gradient Methods,” *Technical Research Report, No. (1), School of Computer Studies Leads University, U.K*, 1986.

[2] M. S. Jameel, G. T. Basheer, A. Y. Al-Bayati, and Z. Y. Algamal, “Parameter estimation of a truncated regression model based on improving numerical optimization algorithms,” in *Journal of Physics: Conference Series*, IOP Publishing, 2021, p. 012059.

[3] N. Andrei, *Nonlinear conjugate gradient methods for unconstrained optimization*. Springer, 2020.

[4] M. Jameel, A. Al-Bayati, and Z. Algamal, “Scaled multi-parameter (SMP) nonlinear QN-algorithms,” in *AIP Conference Proceedings*, AIP Publishing, 2023.

[5] M. S. Jameel, Z. M. Abdullah, F. A. Fawzi, and B. A. Hassan, “A New Shifted Conjugate Gradient

Method" Based" on Shifted Quasi-Newton Condition," in *Journal of Physics: Conference Series*, IOP Publishing, 2021, p. 12105.

[6] X.-S. Yang, *Optimization techniques and applications with examples*. John Wiley & Sons, 2018.

[7] V. Choudhary, K. Mehta, S. Desai, A. Nair, R. Iyer, and M. Gupta, "Advancements in Quasi-Newton Methods for Large-Scale Optimization".

[8] S. Nakayama, Y. Narushima, and H. Yabe, "Memoryless quasi-Newton methods based on spectral-scaling Broyden family for unconstrained optimization.," *Journal of Industrial & Management Optimization*, vol. 15, no. 4, 2019.

[9] Y. H. Dai and Y. Yuan, "A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property," *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 177–182, 1999, doi: 10.1137/S1052623497318992.

[10] H. Tomizuka and H. Yabe, "A Hybrid Conjugate Gradient method for unconstrained Optimization," 2004.

[11] Y. Dai and L. Liao, "New conjugacy conditions and related non-linear CG methods' Appl," *Math optim.,(43). Spring-verlag, New York,* 2001.

[12] M. Al-Baali, "Descent property and global convergence of the Fletcher—Reeves method with inexact line search," *IMA Journal of Numerical Analysis*, vol. 5, no. 1, pp. 121–124, 1985.

[13] L. Zhang and W. Zhou, "Two descent hybrid conjugate gradient methods for optimization," *J Comput Appl Math*, vol. 216, no. 1, pp. 251–264, 2008.