Journal of Applied Analysis and Computation
Volume 14, Number 1, February 2024, 458–472
Website:http://www.jaac-online.com
DOI:10.11948/20230268

DEVELOPING A NEW CONJUGATE GRADIENT ALGORITHM WITH THE BENEFIT OF SOME DESIRABLE PROPERTIES OF THE NEWTON ALGORITHM FOR UNCONSTRAINED OPTIMIZATION

Naima Hamel[1,†], Noureddine Benrabia[2], Mourad Ghiat[1] and Hamza Guebbai[1]

Abstract The conjugate gradient method and the Newton method are both numerical optimization techniques. In this paper, we aim to combine some desirable characteristics of these two methods while avoiding their drawbacks, more specifically, we aim to develop a new optimization algorithm that preserves some essential features of the conjugate gradient algorithm, including the simplicity, the low memory requirements, the ability to solve large scale problems and the convergence to the solution regardless of the starting vector (global convergence). At the same time, this new algorithm approches the quadratic convergence behavior of the Newton method in the numerical sense while avoiding the computational cost of evaluating the Hessian matrix directly and the sensitivity of the selected starting vector. To do this, we propose a new hybrid conjugate gradient method by linking (CD) and (WYL) methods in a convex blend, the hybridization paramater is computed so that the new search direction accords with the Newton direction, but avoids the computational cost of evaluating the Hessian matrix directly by using the secant equation. This makes the proposed algorithm useful for solving large scale optimization problems. The sufficient descent condition is verified, also the global convergence is proved under a strong Wolfe Powel line search. The numerical tests show that, the proposed algorithm provides the quadratic convergence behavior and confirm its efficiency as it outperformed both the (WYL) and (CD) algorithms.

Keywords Uconstraind optimization, conjugate gradient algorithm, newton method, quadratic convergence behavior, global convergence.

MSC(2010) 90C06, 90C26, 49M15, 90C30, 65K05.

[†]The corresponding author.
[1]Laboratoire de Mathématiques Appliquées et de Modélisation, Université 8 Mai 1945 Guelma, B.P. 401, 24000 Guelma, Algeria
[2]Département de Mathématiques et Informatique, Université Mohamed-Chérif Messaadia, B.P. 1553, 41000 Souk Ahras, Algeria
Email: hamel.naima@univ-guelma.dz & hamelnaima24@gmail.com(N. Hamel), n.benrabia@univ-soukahras.dz & noureddinebenrabia@yahoo.fr(N. Benrabia), ghiat.mourad@univ-guelma.dz & mourad.ghi24@gmail.com(M. Ghiat), guebbai.hamza@univ-guelma.dz & guebaihamza@yahoo.fr(H. Guebbai)

1. Introduction

In this paper, we are interested in the minimization of a function with n variables, $n \in N^*$. Consider the nonlinear unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{1.1}$$

where, f: $\mathbb{R}^n \mapsto \mathbb{R}$ is a smooth nonlinear function and its gradient is available and denoted by $g = \nabla f(x)$. Mathematicians have developed many numerical techniques to solve (1.1), among which the steepest descent methods (see e.g. [21, 30]), the Newton methods (see e.g. [24, 25]) conjugate gradient methods (see e.g. [2, 19, 28]) and quasi-Newton methods (see e.g. [26, 27]).

The basis of all these methods is to start with an appropriate initial vector $x_0 \in \mathbb{R}^n$ and generate a sequence $\{x_k\}_{k \geq 0}$, as follows

$$x_{k+1} = x_k + \alpha_k d_k, \qquad k \geq 0, \tag{1.2}$$

where, α_k is the step size determined using a line search technique, and d_k is the search direction that identifies the various methods to solve the problem (1.1). In this work we focus on the Newton method and the conjugate gradient method.

The search direction of the Newton method is calculated as follows

$$d_{k+1} = -\nabla^2 f(x_{k+1})^{-1} g_{k+1}, \tag{1.3}$$

where, $\nabla^2 f(x_{k+1})$ is the Hessian matrix of f. When initialized point near the solution, the Newton method provides a quadratic convergence rate because it uses the second derivative information to generate the search direction. However, the Newton method is efficient for small and medium-sized problems and is not suitable for large scale problems in terms of the storage and the computational cost of evaluating the Hessian matrix [20].

The conjugate gradient method is much more useful and practical for solving (1.1), especially for large-scale cases, due to its simplicity, low memory requirements as it only uses the first derivative information [20]. The conjugate gradient method has the global convergence property, which allows it to converge to the optimal solution regardless of the starting vector selected. The search direction given as

$$d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_k d_k, \tag{1.4}$$

depending on the choice of the parameter $\beta_k \in \mathbb{R}$ known as the conjugate gradient parameter, there are several different conjugate gradient algorithms. In the following we are going to mention some famous formulas for this parameter.

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \quad \text{(HS - Hestenes and Stiefel [15])},$$

$$\beta_k^{FR} = \frac{\| g_{k+1} \|^2}{\| g_k \|^2}, \quad \text{(FR - Fletcher and Reeves [13])},$$

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\| g_k \|^2}, \quad \text{(PRP - Polak and Ribére [22, 23])},$$

$$\beta_k^{CD} = \frac{\| g_{k+1} \|^2}{-d_k^T g_k}, \quad \text{(CD - conjugate descent [12])},$$

$$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k}, \quad \text{(LS - Liu and Storey [18])},$$

$$\beta_k^{DY} = \frac{\| g_{k+1} \|^2}{d_k^T y_k}, \quad \text{(DY - Dai and Yuan [8])},$$

$$\beta_k^{WYL} = \frac{g_{k+1}^T (g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k)}{\| g_k \|^2}, \quad \text{(WYL- Wei, Yao and Liu [16,29])}.$$

Where, $y_k = g_{k+1} - g_k$.

Conjugate gradient algorithms are classified into three major categories:classical methods, modified methods and hybrid methods.

The methods (HS),(FR), (PRP), (CD), (LS), (DY) are known as classical methods due to their simplicity.

The (WYL) conjugate gradient method was proposed by Wei [16,29] as a modified version of the PRP classical method in order to improve it and make it more efficient. This method not only has nice numerical experiments but also satisfies the sufficient descent condition and has global convergence properties.

Hybrid conjugate gradient methods are based on combining the classical or the modified methods in order to build new practical ones that have the advantages of the methods to be combined. So, several hybrid methods are suggested, for example, Andrei [6] proposed combining the (DY) and (HS) conjugate gradient methods as a convex combination and distinguished this method by making its search direction is the Newtonian direction using the secant equation to avoid the evaluation of the Hessien matrix. Motivated by Andrei's idea [6], recently Fanar and Ghada [11], Abdullah and Jamalaldeen [1] and Djordjević [9] derived new hybrid conjugate gradient methods satisfy the sufficient descent property in such a way that Newton directions are employed. Abubakar etc [3,4] proposed a combination between two conjugate gradient methods in such a way that the new search direction is close to the direction of the memoryless Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton approach.

Here, in this paper, inspired by Andrei's work [6], we propose a new hybrid conjugate gradient algorithm that links (WYL) and (CD) methods based on the Newton direction in order to gain some desirable properties of both conjugate gradient and Newtonian methods while avoiding their disadvantages. more specifically, our focus is to preserve the essential features of the conjugate gradient algorithm, including its simplicity, ability to solve large scale problems, and global convergence property, which allows the algorithm to converge to the optimal solution whatever the starting vector selected. Additionally, we aim to retain the fast quadratic convergence behavior of the Newton method in the numerical sense, while avoiding the expensive computation of the Hessian matrix directly and the sensitivity of its convergence to the starting vector selected.

Since the proposed method is constructed with the aim of approaching the quadratic convergence behavior of the Newton method, we propose a new numerical test based on some test functions chosen from [7] with different dimensions and following the next steps

- For each iteration we calculate the error ratios r_k with two successive iterations

$$r_k = \frac{\| x_{k+1} - x^* \|}{\| x_k - x^* \|^2} \tag{1.5}$$

where, x^* is the exacte solution of the problem.

- Plotting $log(r_k)$ versus k and showing that r_k tends to converge towards a constant value, i.e. r_k approches a constant value which proved that, the proposed method provides a quadratic convergence behavior in the numerical sense.

2. Hybrid conjugate gradient algorithm

In this section, we describe our new conjugate gradient method for large scale unconstrained optimization problems, computing the parameter β_k, denoted in this paper by β_k^{wylcd}, as a convex combination between β_k^{CD} and β_k^{WYL}, i.e.

$$\beta_k^{wylcd} = (1-\gamma_k)\beta_k^{WYL} + \gamma_k \beta_k^{CD}, \tag{2.1}$$

where, $\gamma_k \in [0,1]$.

So the direction d_k, is given by

$$d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + (1-\gamma_k)\beta_k^{WYL} d_k + \gamma_k \beta_k^{CD} d_k. \tag{2.2}$$

If $\gamma_k = 0$, then $\beta_k^{wylcd} = \beta_k^{WYL}$ and if $\gamma_k = 1$, then $\beta_k^{wylcd} = \beta_k^{CD}$. On the other hand if $0 < \gamma_k < 1$ then β_k^{wylcd} is the convex combination between β_k^{CD} and β_k^{WYL}.

Assume that $\nabla^2 f(x_k)^{-1}$ exists for all $k \geq 0$ for the objective function f.

As we know, the Newton method has quadratic convergence properties, so we are going to build a new hybrid conjugate gradient method accords with the Newton method. To do this, motivated by Andrei's work [6] we compute the γ_k in such a manner that our search direction given by the relation (2.2) is equal to the Newton direction, i.e.

$$-g_{k+1} + (1-\gamma_k)\beta_k^{WYL} d_k + \gamma_k \beta_k^{CD} d_k = -\nabla^2 f(x_{k+1})^{-1} g_{k+1}. \tag{2.3}$$

Multiplying both sides of the equation (2.3) by $s_k^T \nabla^2 f(x_{k+1})$ from the left we obtain

$$\begin{aligned} &- s_k^T \nabla^2 f(x_{k+1}) g_{k+1} + (1-\gamma_k)\beta_k^{WYL} s_k^T \nabla^2 f(x_{k+1}) d_k + \gamma_k \beta_k^{CD} s_k^T \nabla^2 f(x_{k+1}) d_k \\ &= - s_k^T g_{k+1}, \end{aligned} \tag{2.4}$$

where $s_k = x_{k+1} - x_k$.

Following some algebraic computations, we arrive at

$$\gamma_k = \frac{-s_k^T g_{k+1} + s_k^T \nabla^2 f(x_{k+1}) g_{k+1} - \beta_k^{WYL} s_k^T \nabla^2 f(x_{k+1}) d_k}{(-\beta_k^{WYL} + \beta_k^{CD}) s_k^T \nabla^2 f(x_{k+1}) d_k}. \tag{2.5}$$

To compute γ_k, we have to get the Hessian matrix of the objective function, but we know that for large-scale problems, computing the Hessian matrix is either impossible or expensive in practice. knowing that, for quazi Newton algorithms the approximation matrix B_k to the Hessien matrix $\nabla^2 f(x_k)$ is updated so that the new matrix B_{k+1} satisfies the secant equation $B_{k+1}s_k = y_k$. So, to obtain a widely used problem solving algorithm, we assume that the pair (s_k, y_k) satisfies the secant equation

$$\nabla^2 f(x_{k+1}) s_k = y_k,$$

i.e.

$$s_k^T \nabla^2 f(x_{k+1}) = y_k^T,$$

therefore, we obtain

$$\gamma_k = \frac{-s_k^T g_{k+1} + y_k^T g_{k+1} - \beta_k^{WYL} y_k^T d_k}{(-\beta_k^{WYL} + \beta_k^{CD}) y_k^T d_k}. \tag{2.6}$$

Clearly, we have constructed a new hybrid conjugate gradient method accords with the Newton method, but the iterative process is simple and it is designed to solve large-scale problems because we have avoided the computational cost of evaluating the Hessian matrix directly by using the secant equation.

Now, we describe the proposed algorithm named "wylcd algorithm" which has some good characteristics of both conjugate gradient algorithm and the Newton algorithm.

2.1. Wylcd Algorithm

Step 0. Choose the initial point $x_0 \in \mathbb{R}^n$, $\epsilon > 0$.
Calculate $f_0 = f(x_0)$ and $g_0 = \nabla f(x_0)$.
Set $d_0 = -g_0$, the initial guess $\alpha_0 = \frac{1}{\|g_0\|}$. Let $k = 0$.
Step 1. Test a criterion to stop the iterations, i.e. if $\| g_k \| \leq \epsilon$ then stop. Otherwise go to step 2.
Step 2. Compute the step size α_k using the strong Wolfe Powell line search

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \tag{2.7}$$

$$|g_{k+1}^T d_k| \leq \sigma |g_k^T d_k|. \tag{2.8}$$

Where, $0 < \delta < \frac{1}{2}, 0 < \sigma < \frac{1}{5}$.
Step 3. Updating the next iterate by $x_{k+1} = x_k + \alpha_k d_k$. Compute $g_{k+1} = \nabla f(x_{k+1})$, $y_k = g_{k+1} - g_k$ and $s_k = x_{k+1} - x_k$.
Step 4. If $(-\beta_k^{WYL} + \beta_k^{CD}) y_k^T d_k = 0$, then $\gamma_k = 0$, else calculate γ_k as in (2.6).
Step 5. If $\gamma_k \leq 0$, then calculate $\beta_k^{wylcd} = \beta_k^{WYL}$.
If $\gamma_k \geq 1$, then calculate $\beta_k^{wylcd} = \beta_k^{CD}$.
If $0 < \gamma_k < 1$, then calculate β_k^{wylcd} as in (2.1).
Step 6. Compute $d_{k+1} = -g_{k+1} + \beta_k^{wylcd} d_k$. Set the initial guess $\alpha_k = \alpha_{k-1} \frac{\|d_{k-1}\|}{\|d_k\|}$.
Step 7. Let $k = k + 1$, go to step 1.

3. The sufficient descent condition

Theorem 3.1. *Suppose that the sequences $\{g_k\}_{k\geq 0}$ and $\{d_k\}_{k\geq 0}$ are generated by "wylcd Algorithm", assume that α_k is determined by the strong Wolfe powel line search* (2.7) *and* (2.8)*, if $0 < \sigma < \frac{1}{5}$, then the sufficient descent condition*

$$g_{k+1}^T d_{k+1} \leq -c \| g_{k+1} \|^2 \tag{3.1}$$

holds.

Proof. If $\gamma_k = 0$, then $\beta_k^{wylcd} = \beta_k^{WYL}$ and if $\gamma_k = 1$, then $\beta_k^{wylcd} = \beta_k^{CD}$, the sufficient decent condition is allready proven in [16,29] and [12] respectively.

Now, we poove the sufficient descent condition in the case $0 < \gamma_k < 1$.

From (2.1)

$$\begin{aligned} &|\beta_k^{wylcd}| \leq |\beta_k^{WYL}| + |\beta_k^{CD}| \\ \leq &\frac{\| g_{k+1} \|^2 + \frac{\|g_{k+1}\|}{\|g_k\|} \| g_{k+1} \|\| g_k \|}{\| g_k \|^2} + \frac{\| g_{k+1} \|^2}{| - g_k^T d_k|} \\ = &\frac{2 \| g_{k+1} \|^2}{\| g_k \|^2} + \frac{\| g_{k+1} \|^2}{| - g_k^T d_k|}. \end{aligned}$$

Then, from the above inequality we obtain

$$|\beta_k^{wyld} g_{k+1}^T d_k| \leq \frac{2 \| g_{k+1} \|^2}{\| g_k \|^2} |g_{k+1}^T d_k| + \frac{\| g_{k+1} \|^2}{| - g_k^T d_k|} |g_{k+1}^T d_k|, \tag{3.2}$$

using (2.8) we get

$$|\beta_k^{wylcd} g_{k+1}^T d_k| \leq \frac{2 \| g_{k+1} \|^2}{\| g_k \|^2} \sigma |g_k^T d_k| + \sigma \| g_{k+1} \|^2, \tag{3.3}$$

multiplying (2.2) by g_{k+1} we find

$$g_{k+1}^T d_{k+1} = - \| g_{k+1} \|^2 + \beta_k^{wylcd} g_{k+1}^T d_k. \tag{3.4}$$

Then

$$\frac{g_{k+1}^T d_{k+1}}{\| g_{k+1} \|^2} = -1 + \beta_k^{wylcd} \frac{g_{k+1}^T d_k}{\| g_{k+1} \|^2}. \tag{3.5}$$

Now, let's prove the descent property of the direction d_k by induction. We have $g_1^T d_1 = - \| g_1 \|^2$, we suppose that d_i, $i = 1, 2, \ldots, k$ are all descent directions ($g_i^T d_i < 0$).

From (3.3), it results

$$|\beta_k^{wylcd} g_{k+1}^T d_k| \leq -2\sigma \| g_{k+1} \|^2 \frac{g_k^T d_k}{\| g_k \|^2} + \sigma \| g_{k+1} \|^2 . \tag{3.6}$$

Then

$$\begin{aligned} &2\sigma \| g_{k+1} \|^2 \frac{g_k^T d_k}{\| g_k \|^2} - \sigma \| g_{k+1} \|^2 \\ \leq &\beta_k^{wylcd} g_{k+1}^T d_k \\ \leq &- 2\sigma \| g_{k+1} \|^2 \frac{g_k^T d_k}{\| g_k \|^2} + \sigma \| g_{k+1} \|^2, \end{aligned} \tag{3.7}$$

from (3.5) and (3.7) we deduce

$$\frac{g_{k+1}^T d_{k+1}}{\| g_{k+1} \|^2} \leq -1 - 2\sigma \frac{g_k^T d_k}{\| g_k \|^2} + \sigma. \tag{3.8}$$

Repeating this process

$$
\begin{aligned}
&\frac{g_{k+1}^T d_{k+1}}{\| g_{k+1} \|^2} \\
&\leq -1 - 2\sigma \frac{g_k^T d_k}{\| g_k \|^2} + \sigma \\
&\leq -1 - 2\sigma(-1 - 2\sigma \frac{g_{k-1}^T d_{k-1}}{\| g_{k-1} \|^2} + \sigma) + \sigma \\
&\leq -1 + 2\sigma + (2\sigma)^2(-1 - 2\sigma \frac{g_{k-2}^T d_{k-2}}{\| g_{k-2} \|^2} + \sigma) - 2\sigma^2 + \sigma \\
&\leq -1 + 2\sigma - (2\sigma)^2 - (2\sigma)^3(-1 - 2\sigma \frac{g_{k-3}^T d_{k-3}}{\| g_{k-3} \|^2} + \sigma) + \sigma(2\sigma)^2 - 2\sigma^2 + \sigma \\
&\leq -1 + 2\sigma - (2\sigma)^2 + (2\sigma)^3 + (2\sigma)^4(-1 - 2\sigma \frac{g_{k-4}^T d_{k-4}}{\| g_{k-4} \|^2} + \sigma) - \sigma(2\sigma)^3 + \sigma(2\sigma)^2 \\
&\quad -2\sigma^2 + \sigma \\
&\vdots \\
&\leq -1 + 2\sigma - (2\sigma)^2 + (2\sigma)^3 - (2\sigma)^4 + \dots \\
&\quad -(2\sigma)^{k-1}(-1 - 2\sigma \frac{g_{k-(k-1)}^T d_{k-(k-1)}}{\| g_{k-(k-1)} \|^2} + \sigma) - \sigma(2\sigma)^3 + \sigma(2\sigma)^2 - 2\sigma^2 + \sigma \\
&\leq -1 + 2\sigma - (2\sigma)^2 + (2\sigma)^3 - (2\sigma)^4 + \dots + (2\sigma)^{k-1} + (2\sigma)^k \frac{g_1^T d_1}{\| g_1 \|^2} - \sigma(2\sigma)^{k-1} \\
&\quad -\sigma(2\sigma)^3 + \sigma(2\sigma)^2 - 2\sigma^2 + \sigma, \qquad (3.9)
\end{aligned}
$$

using $g_1^T d_1 = - \| g_1 \|^2$, the inequality (3.9) becomes

$$
\begin{aligned}
\frac{g_{k+1}^T d_{k+1}}{\| g_{k+1} \|^2} &\leq -1 + [(2\sigma + (2\sigma)^3 + \dots + (2\sigma)^{k-1}) - ((2\sigma)^2 + (2\sigma)^4 + \dots + (2\sigma)^k)] \\
&\quad +[(\sigma + \sigma(2\sigma)^2 + \dots + \sigma(2\sigma)^{k-2}) - (2\sigma^2 + \sigma(2\sigma)^3 + \dots + \sigma(2\sigma)^{k-1})] \\
&\leq -1 + [(2\sigma + (2\sigma)^3 + \dots + (2\sigma)^{k-1}) + ((2\sigma)^2 + (2\sigma)^4 + \dots + (2\sigma)^k)] \\
&\quad +[(\sigma + \sigma(2\sigma)^2 + \dots + \sigma(2\sigma)^{k-2}) + (2\sigma^2 + \sigma(2\sigma)^3 + \dots + \sigma(2\sigma)^{k-1})] \\
&= -1 + \sum_{j=1}^{k} (2\sigma)^j + \sigma \sum_{j=0}^{k-1} (2\sigma)^j \\
&= -2 + \sum_{j=0}^{k} (2\sigma)^j + \sigma \sum_{j=0}^{k-1} (2\sigma)^j, \qquad (3.10)
\end{aligned}
$$

we have

$$
\sum_{j=0}^{k} (2\sigma)^j < \sum_{j=0}^{\infty} (2\sigma)^j = \frac{1}{1-(2\sigma)} \qquad \text{and} \qquad \sum_{j=0}^{k-1} (2\sigma)^j < \sum_{j=0}^{\infty} (2\sigma)^j = \frac{1}{1-(2\sigma)}.
$$

Therefore, the inequality (3.10) becomes

$$
\frac{g_{k+1}^T d_{k+1}}{\| g_{k+1} \|^2} \leq -(2 - \frac{1+\sigma}{1-(2\sigma)}), \qquad (3.11)
$$

taking $0 < \sigma < \frac{1}{5}$, then

$$g_{k+1}^T d_{k+1} \leq -(2 - \frac{1+\sigma}{1-(2\sigma)}) \parallel g_{k+1} \parallel^2 < 0. \tag{3.12}$$

So, by induction $g_k^T d_k < 0$ holds for all $k \geq 0$.

Now, we prove the sufficient descent condition of d_k. If $0 < \sigma < \frac{1}{5}$, it suffices to take $c = 2 - \frac{1+\sigma}{1-(2\sigma)}$, where $0 < c < 1$. Then from the inequality (3.12), it results

$$g_{k+1}^T d_{k+1} \leq -c \parallel g_{k+1} \parallel^2, \tag{3.13}$$

which indicates that the sufficient descent condition holds. □

4. Convergence properties

The following assumptions on the objective function f are required to establish the global convergence of the proposed algorithm.

Assumption 1 ([5]). H1. *The level set $\mathcal{H} = \{x \in R^n / f(x) \leq f(x_0)\}$ is bounded where x_0 is the initial vector.*

H2. *In some neighborhood $\mathcal{Q}$ of $\mathcal{H}$, the function f is continuously differentiable and its gradient is Lipschitz continuous, i.e. $\exists l > 0$ such that*

$$\parallel g(x) - g(y) \parallel \leq l \parallel x - y \parallel \qquad \text{for all} \quad x, y \in \mathcal{Q}. \tag{4.1}$$

These hypotheses imply that $\exists \bar{r} > 0$ such that

$$\parallel g(x) \parallel \leq \bar{r}, \forall x \in \mathcal{H}. \tag{4.2}$$

Lemma 1 ([8]). *Let the above assumptions H1 and H2 hold and consider the methods formulated by (1.2), (1.4), where $\{d_k\}$ is a descent direction, and α_k is calculatd by the strong Wolfe Powel line search. If*

$$\sum_{k \geq 0} \frac{1}{\parallel d_k \parallel} = +\infty \tag{4.3}$$

then

$$\lim_{k \mapsto \infty} inf \parallel g_k \parallel = 0. \tag{4.4}$$

Lemma 2 ([17]). *Suppose that the above assumptions H1 and H2 hold. If d_k is a descent direction and the step length α_k satisfies*

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k, \quad \sigma < 1, \tag{4.5}$$

then

$$\alpha_k \geq \frac{1-\sigma}{l} \frac{|d_k^T g_k|}{\parallel d_k \parallel^2}. \tag{4.6}$$

Proof. From (4.5), (4.1) it results that:

$$\begin{aligned} & -(1-\sigma) g_k^T d_k \\ \leq & \sigma g_k^T d_k - g_k^T d_k \end{aligned}$$

$$\begin{aligned}
&\leq d_k^T (g_{k+1} - g_k)\\
&= d_k^T y_k\\
&\leq \| d_k \| \| y_k \|\\
&\leq l\alpha_k \| d_k \|^2 .
\end{aligned}$$

Then, (4.6) holds. □

According to the Lemma 2, conditions (2.8), (3.1) and (4.2) we deduce that α_k obtained in the new wylcd algorithm is not equal to 0, i.e. $\exists p > 0$ such that

$$\alpha_k \geq p, \quad \text{for all} \quad k \geq 0. \tag{4.7}$$

Theorem 4.1. *Suppose that Assumption 1 holds. Let the sequence $\{x_k\}_{k\geq 0}$ be generated by the proposed "wylcd Algorithm". Then*

$$\lim_{k\mapsto\infty} inf \| g_k \| = 0. \tag{4.8}$$

Proof. Suppose that (4.8) is false. Then there exists $r > 0$ such that

$$\| g_k \| > r. \tag{4.9}$$

From (4.9) and (3.1) we have

$$-g_k^T d_k \geq c \| g_k \|^2 \geq cr^2, \tag{4.10}$$

$$\frac{1}{-g_k^T d_k} \leq \frac{1}{cr^2}. \tag{4.11}$$

Knowing that $s_k = x_{k+1} - x_k$, let $A = \max\{\| x - y \| / x, y \in \mathcal{H}\}$ be the diameter of the level set $\mathcal{H}$.i.e $\| s_k \| \leq A$.

From (2.1) and (2.2), we have

$$\| d_{k+1} \| \leq \| g_{k+1} \| + |\beta_k^{wylcd}| \| d_k \| . \tag{4.12}$$

Concerning the boundedness of the β_k^{wylcd} there are three cases. If $\gamma_k = 0$, then $\beta_k^{wylcd} = \beta_k^{WYL}$ and if $\gamma_k = 1$, then $\beta_k^{wylcd} = \beta_k^{CD}$, these two cases are allready proven in [16, 29] and [12] respectively.

Now, if $0 < \gamma_k < 1$

$$\begin{aligned}
|\beta_k^{wylcd}| &= |(1 - \gamma_k)\beta_k^{WYL} + \gamma_k \beta_k^{CD}|\\
&\leq |\beta_k^{WYL}| + |\beta_k^{CD}|\\
&\leq \frac{\| g_{k+1} \|^2 + \frac{\|g_{k+1}\|}{\|g_k\|} \| g_{k+1} \| \| g_k \|}{\| g_k \|^2} + \frac{\| g_{k+1} \|^2}{|-g_k^T d_k|}\\
&= \frac{2 \| g_{k+1} \|^2}{\| g_k \|^2} + \frac{\| g_{k+1} \|^2}{|-g_k^T d_k|}.
\end{aligned} \tag{4.13}$$

From (4.9), (4.11) and(4.2) it results

$$|\beta_k^{wylcd}| \leq \tfrac{2\bar{r}^2}{r^2} + \tfrac{\bar{r}^2}{cr^2} = E. \tag{4.14}$$

According to (4.14) and (4.2), the relation (4.12) becomes

$$\| d_{k+1} \| \leq \bar{r} + E \| d_k \| . \tag{4.15}$$

Using $\| d_k \| = \frac{\| s_k \|}{\alpha_k}$ and from (4.7), we get

$$\| d_{k+1} \| \leq \bar{r} + E \frac{\| s_k \|}{\alpha_k} \leq \bar{r} + E \frac{A}{p}.$$

Therefore

$$\sum_{k \geq 0} \frac{1}{\| d_k \|} = +\infty. \tag{4.16}$$

Now, applying Lemma 1, we conclude that

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

this is a contradiction with (4.9), so we have proved (4.8). □

5. Numerical experiments

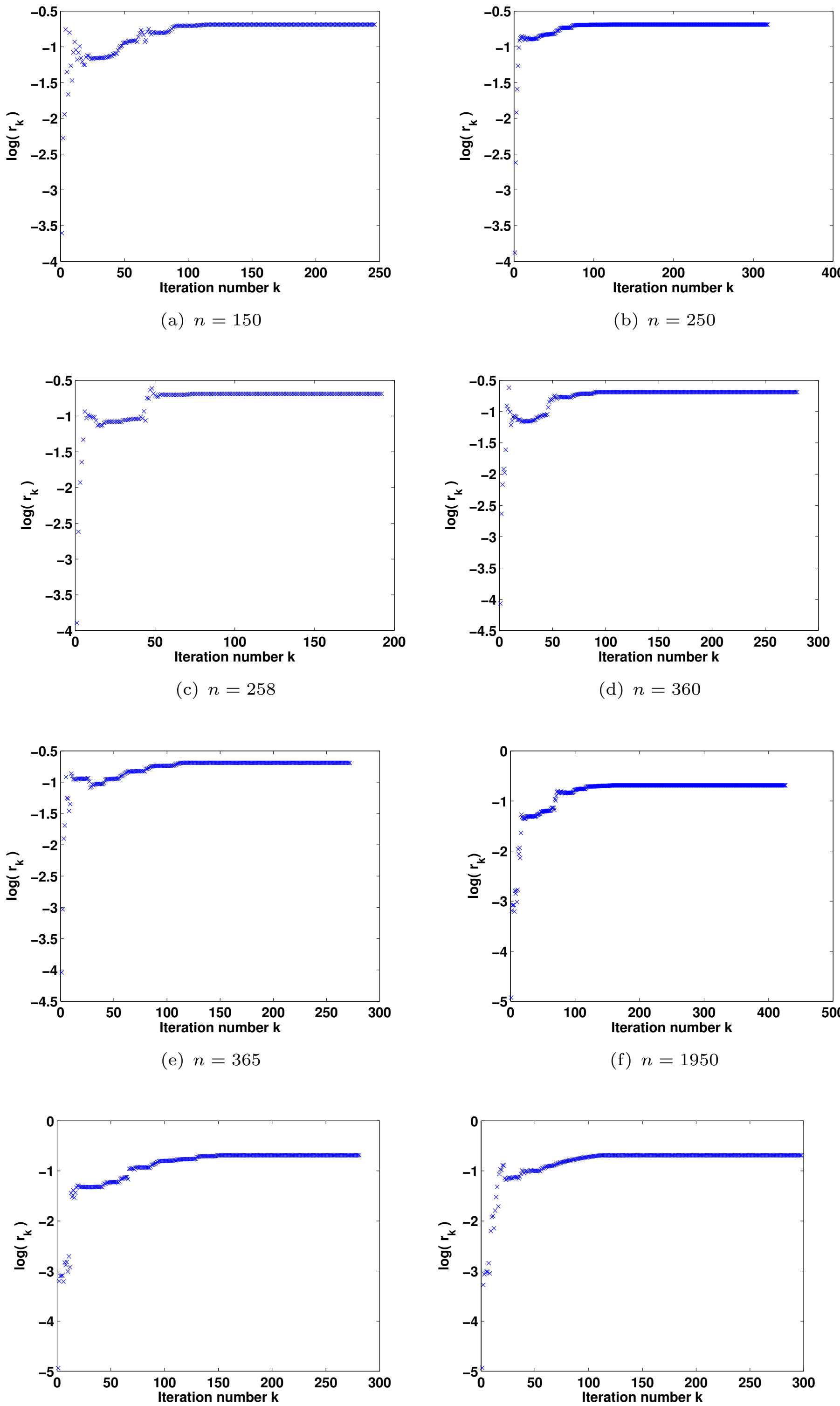
In this section, we are going to describe the numerical results of the new proposed algorithm (wylcd algorithm). Firstly, we prove numerically that, the proposed algorithm has the quadratic convergence behavior, secondly, we analyze the efficiency of the proposed algorithm by comparing it to the (WYL) and (CD) methods. In the next numerical experiments, all codes are compiled with a PC with the following specifications: Intel(R) Core(TM) i5-3210M CPU 2.50GHz 2.50 GHz, 4.00 Go RAM, using the profile of Dolan and Moré [10] as an evaluation tool. All algorithms employ the strong Wolfe Powel line search conditions with the parameters $\delta = 0.0001$ and $\sigma = 0.1$ and terminate when $\| g_k \|_\infty \leq 10^{-6}$.

The quadratic convergence behavior of the proposed algorithm

As we say in section 2, the Newton method has a quadratic convergence rate and this property is very favorable because it means a fast convergence to the optimal solution, so we have constructed a new hybrid conjugate gradient algorithm based on the Newton direction, with the aim of obtaining the fast convergence behavior of the Newton method in the numerical sense and achieving some essential features of the conjugate gradient algorithm. In the following, we propose a new numerical test which demonstrates the quadratic convergence behavior of the proposed algorithm using some test problems chosen from [7] with dimensions n: $150, 250, 258, 360, 365, 1950, 2000, 2005$ and for each iteration we calculate the error ratios r_k with two successive iterations

$$r_k = \frac{\| x_{k+1} - x^* \|}{\| x_k - x^* \|^2} \tag{5.1}$$

where, x^* is the exacte solution of the selected problem. We plot $log(r_k)$ versus k (see figure 1).



(a) $n = 150$

(b) $n = 250$

(c) $n = 258$

(d) $n = 360$

(e) $n = 365$

(f) $n = 1950$

(g) $n = 2000$

(h) $n = 2005$

Figure 1. The quadratic convergence behavior of the proposed algorithm

As shown by figure 1, $log(r_k) < 0$ i.e. $r_k < 1$. Then, we can observe that r_k tends to converge to a constant value i.e. r_k approches a constant value. This observation proves the quadratic convergence behavior of the new algorithm in the numerical sense.

Numerical comparisons

For the evaluation of the effectiveness of our "wylcd algorithm", we tested it against the (WYL) [16,29] and (CD) [12] algorithms from which it was built using the same test problems chosen from [7], for each function we performed numerical experiments for the number of variables: 2, 10, ..., 10000. Figure 2, 3 and 4 plot the performance of these algorithms based on these indicators: CPU time, Number of iterations, gradient evaluations.

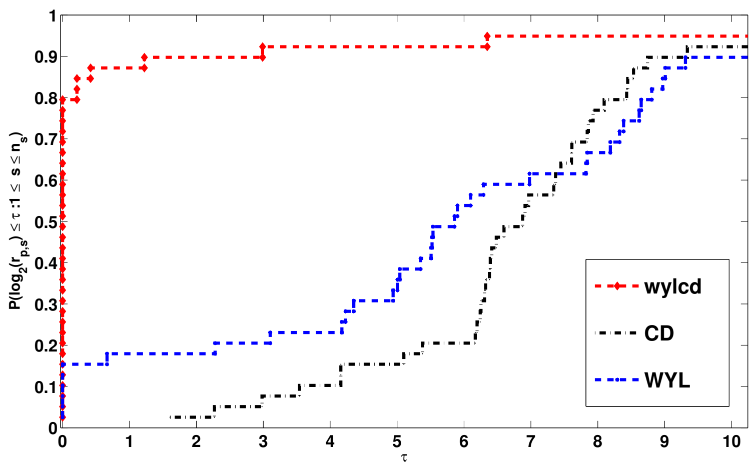


Figure 2. Performance profile for CPU time

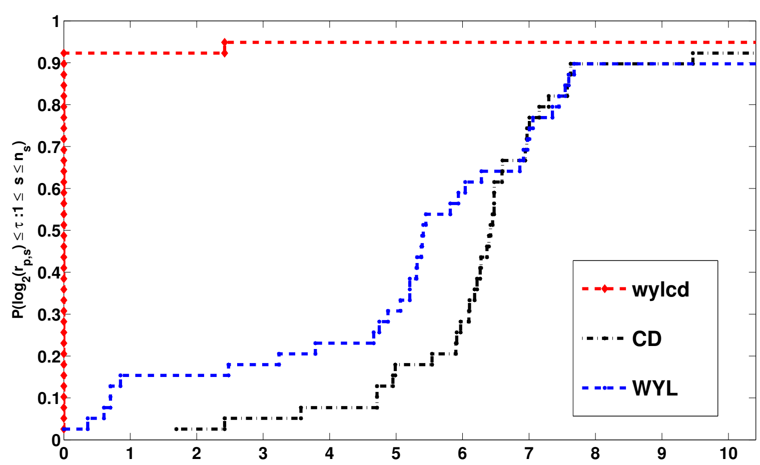


Figure 3. Performance profile for the number of iterations.

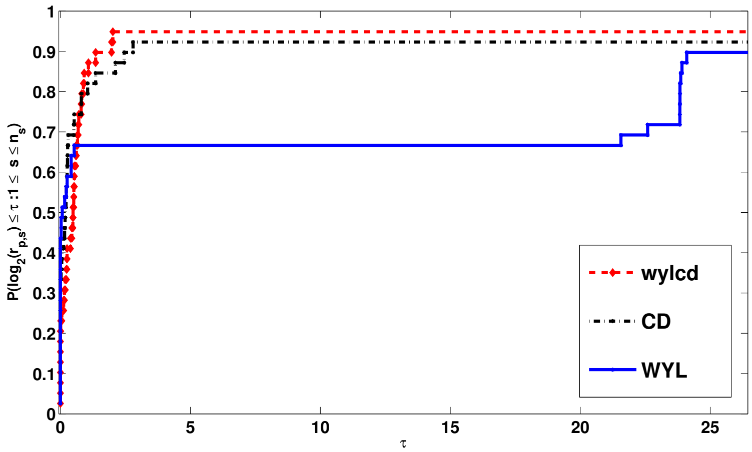


Figure 4. Performance profile for gradient evaluations.

As all the above figures show, the new wylcd algorithm is clearly superior to the other algorithms.

6. Conclusion

In this research paper we have presented a new optimization algorithm that combines some essential features of both the conjugate gradient and Newtonian methods while avoiding their disadvantages. Specifically, we focused on preserving the simplicity, the low memory requirements, the suitability for solving problems when the dimension is large and the global convergence of the conjugate gradient method. Simultaneously, we focused on retaining the fast quadratic convergence behavior of the Newton method in the numerical sense, while avoiding the computational cost of evaluating the Hessian matrix directly and the sensitivity of its convergence to the initial vector chosen. The proposed algorithm is a hybrid conjugate gradient method that blends the (CD) and (WYL) methods in a convex manner. This algorithm is constructed to be closely related to the Newton method without needing to evaluate the Hessian matrix due to the use of the secant equation, this makes it useful for solving large scale optimization problems. The sufficient descent condition and global convergence have been proved. The effectiveness of the proposed method has been checked with a set of standard test problems, which showed that the proposed algorithm approches the quadratic convergence behavior of the Newton method and confirmed its superiority over the (WYL) and (CD) algorithms in most cases.

References

[1] Z. M. Abdullah and I. K. Jamalaldeen, *A new hybrid of DY and CGSD conjugate gradient methods for solving unconstrained optimization problems*, Tik. J. of Pure Sci., 2021, 26(5), 86–91.

[2] A. B. Abubakar, et al, *A Liu-Storey-type conjugate gradient method for unconstrained minimization problem with application in motion control*, Journal of King Saud University - Science, 2022, 34(4), 101923.

[3] A. B. Abubakar, P. Kumam, M. Malik, P. Chaipunya and A. H. Ibrahim, *A hybrid FR-DY conjugate gradient algorithm for unconstrained optimization with application in portfolio selection*, AIMS Mathematics, 2021, 6(6), 6506–6527.

[4] A. B. Abubakar, P. Kumam, M. Malik and A. H. Ibrahim, *A hybrid conjugate gradient based approach for solving unconstrained optimization and motion control problems*, Mathematics and Computers in Simulation, 2022, 201, 640–657.

[5] N. Andrei, *A hybrid conjugate gradient algorithm with modified secant condition for unconstrained optimization as a convex combination of Hestenes-Stiefel and Dai-Yuan algorithms*, Studies in Informatics and Control, 2008, 17(4), 373–392.

[6] N. Andrei, *Another hybrid conjugate gradient algorithm for unconstrained optimization*, Numer. Algor., 2008, 47(2), 143–156.

[7] N. Andrei, *An unconstrained optimization test functions*, Collection. Adv. Model. Optim., 2008, 10(1), 147–161.

[8] Y. H. Dai and Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM Journal on optimization, 1999, 10(1), 177–182.

[9] S. S. Djordjević, *New conjugate gradient method as a convex combination of LS and FR methods*, Acta Mathematica Scientia, 2019, 39B(1), 214–228.

[10] E. D. Dolan and J. J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program., 2002, 91(2), 201–213.

[11] N. J. Fanar and M. A. Ghada, *A new hybrid conjugate gradient algorithm as a convex combination of MMWU and RMIL nonlinear problems*, Journal of Interdisciplinary Mathematics, 2021, 24(3), 637–655.

[12] R. Fletcher, *Practical Methods of Optimization. Unconstrained Optimization*, John Wiley and Sons, New York, 1987.

[13] R. Fletcher and C. M. Reeves, *Function minimization by conjugate gradients*, Comput. J., 1964, 7(2), 149–154.

[14] J. C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optimizat., 1992, 2(1), 21–42.

[15] M. R. Hestenes and E. L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Bur. Stand., 1952, 49(6), 409–436.

[16] H. Huang, Z. Wei and S. Yao, *The proof of the sufficient descent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search*, Applied Mathematics and Computation, 2007, 189, 1241–1245.

[17] J. K. Liu and S. J. Li, *New hybrid conjugate gradient method for unconstrained optimization*, Applied Mathematics and Computation, 2014, 245, 36–43.

[18] Y. Liu and C. Storey, *Efficient generalized conjugate gradient algorithms, part 1: Theory*, J. Optim. Theory. Appl., 1991, 69(1), 129–137.

[19] Y. Liu, Z. Zhu and B. Zhang, *Two sufficient descent three-term conjugate gradient methods for unconstrained optimization problems with applications in compressive sensing*, J. Appl. Math. Comput., 2022, 68, 1787–1816.

[20] S. Narayanan and P. Kaelo, *A linear hybridization of Dai-Yuan and Hestenes-Stiefel conjugate gradient method for unconstrained optimization*, Numer. Math. Theor. Meth. Appl., 2021, 14, 527–539.

[21] H. Oviedo, *Implicit steepest descent algorithm for optimization with orthogonality constraints*, Optimization Letters, 2022, 16(6), 1773–1797.

[22] E. Polak and G. Ribiére, *Note sur la convergence de méthodes de directions conjugués*, Rev. Fr. Inf. Rech. Oper., 1969, 3(16), 35–43.

[23] B. T. Polyak, *The conjugate gradient method in extremal problems*, USSR Comput. Math. Math. Phys., 1969, 9(4), 94–112.

[24] B. T. Polyak, *Newton's method and its use in optimization*, European Journal of Operational Research, 2007, 181, 1086–1096.

[25] A. Samson, *Nonlinear Programming: Theories and Algorithms of Some Unconstrained Optimization Methods (Steepest Descent and Newton's Method)*, International Journal of Engineering and Management Research, 2020, 10(2), 1–12.

[26] H. J. M. Shi, Y. Xie, R. Byrd and J. Nocedal, *A noise-tolerant quasi-Newton algorithm for unconstrained optimization*, SIAM Journal on Optimization, 2022, 32(1), 29–55.

[27] Y. Wang, F. Alpak, G. Gao, C. Chen, J. Vink, T. Wells and F. Saaf, *An efficient bi-objective optimization workflow using the distributed quasi-Newton method and its application to well-location optimization*, SPE Journal, 2022, 21(1), 364–380.

[28] H. A. Wasi and M. A. K. Shiker, *Nonlinear conjugate gradient method with modified armijo condition to solve unconstrained optimization*, J. Phys.: Conf. Ser., 2021, 1818, 1–7.

[29] Z. Wei, S. Yao and L. Liu, *The convergence properties of some new conjugate gradient methods*, Applied Mathematics and Computation, 2006, 183, 1341–1350.

[30] X. P. Zhao, J. C. Yao and Y. Yao, *A nonmonotone gradient method for constrained multiobjective optimization problems*, J. Nonlinear Var. Anal., 2022, 6(6), 693–706.