STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **16**, September 2026, pp 2451–2464.
Published online in International Academic Press (www.IAPress.org)

Globally convergent conjugate gradient algorithms for large-scale unconstrained optimization

Mehamdia abd Elhamid [1], Ali Joma Alissa [2,*], Basim A. Hassan [3], Bouaziz Tayeb[4], Ismat Suleiman Salem Abu Sahyoun[5]

[1]*Department of Mathematics, Universite M'Hamed Bougara of Boumerdes, Algeria*
[2] *Education Depertment, Liwa University, Abu Dhabi, United Arab Emirates*
[3] *College of Computers Sciences and Mathematics, university of Mosul, Iraq*
[4]*Department of Mathematics, Universite M'Hamed Bougara of Boumerdes, Algeria*
[5]*Liwa University, Saeed Bin Ahmed Al Otaiba Street (Al Najda Street previously), Al Danah, Baniyas Tower B, Abu Dhabi, United Arab Emirates*

Abstract Optimization methods are widely used to obtain the numerical solution of the optimal control problems arising in scientific and engineering computation, especially for solving large-scale problems. In this paper, based on some modern and computationally efficient methods, two modified conjugate gradient methods (named IHS and IPRP methods) are proposed for unconstrained optimization. Under the strong Wolfe line search (SWLS), the presented methods are proven to be sufficient descent at each iteration. Moreover, we proved that the proposed methods are globally convergent for arbitrary functions and the line search satisfies the strong Wolfe conditions. Numerical tests demonstrate the effectiveness of the IHS and IPRP methods when compared to certain existing methods in view of the Dolan and Moré performance profile.

Keywords Hybrid conjugate gradient method, Inexact line search, Descent condition, Global convergence, Numerical comparisons.

AMS 2010 subject classifications 90C30, 65K05

DOI: 10.19139/soic-2310-5070-3418

1. Introduction

Conjugate gradient (CG) methods are a popular family of iterative algorithms to solve large-scale nonlinear optimization problems due to appropriate features such as no need to calculate the second-order derivatives, low storage and computation, and suitable convergence rate. For more references on advances in the CG method, see [12-22]. Several improved nonlinear conjugate gradient methods and their applications have been extensively investigated in the literature, particularly in nonparametric estimation, image restoration, and regression problems, see [3,4] and [28-33]. In this paper, we consider the following unconstrained optimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$, is continuously differentiable and its gradient is denoted by $g_k = \nabla f(x_k)$.

Conjugate gradient methods are a class of important methods for solving the above problem often using the following iterative rules

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

*Correspondence to: Ali j. Alissa (Email: ali.alissa@lu.ac.ae). Education Depertment, Liwa University, Abu Dhabi, United Arab Emirates).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

where x_k, is the current iteration point, α_k is the steplength computed by certain line search and $d_k \in \mathbb{R}^n$ is defined by

$$d_{k+1} = -g_{k+1} + \beta_k d_k;\ d_0 = -g_0. \tag{1.3}$$

Distinctive CG methods correspond to different choices for the conjugate gradient coefficient β_k, which in turn lead to quite diverse computational efficiency and convergence results of the corresponding methods. Well-known formulate for β_k are called the Fletcher-Reeves [9], Hestenes-Stiefel [24], Polak-Ribière and Polyak [34, 35], Dai-Yuan [5], Conjugate-Descent [10] and Liu-Storey [27] formulate. These parameters are given by the following formulae

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{y_k^T d_k}, \beta_k^{CD} = \frac{\|g_{k+1}\|^2}{-g_k^T d_k},$$

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2},\ \ \beta_k^{HS} = \frac{g_{k+1}^T y_k}{y_k^T d_k},\ \ \beta_k^{LS} = \frac{g_{k+1}^T y_k}{-g_k^T d_k}.$$

Where y_k is defined as the difference between g_{k+1} and g_k, and $\|.\|$ represents the Euclidean norm. In the convergence analysis of CG methods, it is often necessary for the line search to meet the Wolfe (WLS) or strong Wolfe (SWLS) conditions.

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{1.4}$$

and

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k. \tag{1.5}$$

Furthermore, the strong Wolfe (SWLS) conditions encompass (1.4) and

$$\left|g_{k+1}^T d_k\right| \leq -\sigma g_k^T d_k. \tag{1.6}$$

Where $0 < \delta < \sigma < 1$. From a practical computations point of view, if the FR method produces a bad direction and a little step from x_k to x_{k+1}, the next direction and the next step are also probable to be poor unless a reboot along the negative gradient direction is executed [36]. Although there is such a drawback, it has been shown that the FR method has strong convergent properties [25]. The numerical performances of the CD and DY methods are very similar to the FR method since the scalar β_k in these three methods have the same numerator.

In the past few years, the PRP method has generally been regarded to be one of the most efficient CG methods in practical computation. A wonderful property of the PRP method is that it automatically performs a restart if a bad direction occurs [23]. The numerical performances of the HS and LS methods are very similar to the PRP method since the coefficient β_k in these methods have the same numerator. However, the convergence properties of the PRP, HS and LS methods are not so good [37]. In recent years, based on the above six formulas and their hybridization, many works putting effort into seeking new CG methods with not only good convergence properties but also excellent numerical effects were published. In order to achieve global convergence of the PRP method under the inexact linear search, Wei et al. [38] provided a variant of the PRP formula as follows

$$\beta_k^{\text{WYL}} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} g_{k+1}^T g_k}{\|g_k\|^2}.$$

and introduced an algorithm, (named WYL method for short). The WYL method not only preserves the numerical property of the PRP method but also achieves global convergence under the strong Wolfe line search [25]. Yao et al. In reference [39] expanded this concept to the HS method, this modification is referred to as the MHS approach and the parameter β_k is defined within this method as follows

$$\beta_k^{\text{MHS}} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} g_{k+1}^T g_k}{y_k^T d_k}.$$

Based on the strong Wolfe line search, the algorithm generated by [39] is globally convergent and the numerical results have shown that they are promising. Zhang [40] gives two modified CG methods, proposing the following formula

$$\beta_k^{NHS} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|}\left|g_{k+1}^T g_k\right|}{y_k^T d_k}, \quad \beta_k^{NPRP} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|}\left|g_{k+1}^T g_k\right|}{\|g_k\|^2}.$$

The NPRP and NHS methods have sufficient descent conditions and are globally convergent if the SWLS is utilized with the parameter $\sigma < \frac{1}{2}$ [40]. Soon afterward, Huang proposed a variant of the DY method [26], called the MDY method, in which the parameter β_k is yielded by

$$\beta_k^{MDY} = \frac{\|g_{k+1}\|^2 - \frac{g_{k+1}^T d_k}{\|d_k\|^2} g_{k+1}^T d_k}{y_k^T d_k}.$$

The researchers achieved sufficient properties and the global convergence of this method in the context of using SWLS. Numerical results confirm that this method is promising for solving large-scale optimization problems [26]. Recently, Du et al. [8] give two modified CG methods, proposing the following formula

$$\beta_k^{NVHS^*} = \frac{\|g_{k+1}\|^2 - \frac{\left|g_{k+1}^T g_k\right|}{\|g_k\|^2} g_{k+1}^T g_k}{y_k^T d_k}, \text{ and } \beta_k^{NVPRP^*} = \frac{\|g_{k+1}\|^2 - \frac{\left|g_{k+1}^T g_k\right|}{\|g_k\|^2} g_{k+1}^T g_k}{\|g_k\|^2}.$$

If the SWLS is utilized they prove the sufficient descent and the convergence property of these methods for uniformly convex functions [8].

1.1. Motivation and some new formulas

The two methods presented in this work are the result of closely monitoring the construction of conjugate gradient parameters in the existing NHS and NPRP methods. Clearly, β_k^{NHS} and β_k^{NPRP} share the same mathematical expression for the numerator. This common structure suggests a unified framework that can be further developed and enhanced. However, while these methods improve upon their classical counterparts (HS and PRP), they still rely on the absolute value $|g_{k+1}^T g_k|$, which captures only the cosine of the angle between successive gradients, i.e., $\frac{|g_{k+1}^T g_k|}{\|g_{k+1}\|\|g_k\|}$. This term, while informative, ignores the previous search direction d_k, which contains crucial historical information about the search path and the geometry of the objective function. In regions where the objective function exhibits sharp nonlinear behavior, this limitation becomes particularly significant.

This observation constitutes the core motivation for our work: we sought not merely to replace one mathematical term with another, but to reformulate the numerator in a more dynamic and precise manner that better reflects the geometric relationship between vectors. By carefully examining the numerators of the NHS and NPRP methods, we propose that the parameter β_k can be alternatively and more effectively chosen as:

$$\beta_k^{IHS} = \frac{\|g_{k+1}\|^2 - \frac{\theta_k (g_{k+1}^T g_k)^2}{\|d_k\|^2 \|g_{k+1}\|^2}}{y_k^T d_k + \xi_1 \|g_{k+1}\| \|d_k\|}, \quad \theta_k = \frac{\eta_1 (g_{k+1}^T d_k)^2}{\|g_k\|^2}, \quad \eta_1 \in [0,1]. \tag{1.7}$$

That is, we replace the term $\frac{\|g_{k+1}\|}{\|g_k\|}|g_{k+1}^T g_k|$ in β_k^{NHS} with the more adaptive expression $\frac{\theta_k (g_{k+1}^T g_k)^2}{\|d_k\|^2 \|g_{k+1}\|^2}$ in β_k^{IHS}. This transformation introduces an adaptive parameter θ_k that depends on the angle between the current gradient g_{k+1} and the previous search direction d_k, creating a dynamic adjustment mechanism. The term $(g_{k+1}^T g_k)^2$ provides a quadratic, continuously differentiable measure of gradient correlation, in contrast to the absolute value $|g_{k+1}^T g_k|$ which is not differentiable at zero, contributing to more stable numerical behavior. The normalization factor $\|d_k\|^2 \|g_{k+1}\|^2$ ensures scale invariance, a desirable property for conjugate gradient methods. Additionally, we reinforce the denominator by adding a safeguarding term $\xi_1 \|g_{k+1}\| \|d_k\|$, where $\xi_1 > 0$, to ensure numerical stability and facilitate the proof of the sufficient descent property. The parameter $\eta_1 \in [0,1]$ controls the influence of the

adaptive directional term; when $\eta_1 = 0$, the method reduces to a safeguarded version of NHS, while $\eta_1 = 1$ activates the full adaptive term. The parameter $\xi_1 > 0$ serves as a regularization parameter that prevents the denominator from vanishing when $d_k^T(g_{k+1} - g_k)$ is very small. Following the same principle, we define the parameter β_k for the IPRP method as:

$$\beta_k^{IPRP} = \frac{\|g_{k+1}\|^2 - \frac{\theta_k (g_{k+1}^T g_k)^2}{\|d_k\|^2 \|g_{k+1}\|}}{\|g_k\|^2 + \xi_2 \|g_{k+1}\| \|d_k\|}, \quad \theta_k = \frac{\eta_2 (g_{k+1}^T d_k)^2}{\|g_k\|^2}, \quad \eta_2 \in [0,1]. \tag{1.8}$$

Here, $\eta_1, \eta_2 \in [0,1]$ and $\xi_1, \xi_2 > 0$ are constants that control the adaptivity and safeguarding properties of the methods. The denominator in (1.8) incorporates a safeguarding term $\xi_2 \|g_{k+1}\| \|d_k\|$ added to $\|g_k\|^2$, which similarly prevents numerical instability when $\|g_k\|$ becomes very small.

The primary attributes of these methods are as follows:

- Two modified conjugate gradient algorithms, based on the HS and PRP methods, are developed for solving unconstrained optimization problems.
- The search directions generated by the presented methods satisfy the sufficient descent condition at each iteration when combined with a strong Wolfe line search. Furthermore, these modifications are proved to be globally convergent under standard assumptions.
- Based on extensive numerical experiments, the results demonstrate that our methods perform better compared to several existing methods for solving unconstrained optimization problems. Particularly in terms of CPU time, our methods prove superior for the given test problems.
- The proposed methods are successfully applied to solving conditional model regression problems, with lower computational cost, indicating the encouraging applicability of our approaches.

2. The sufficient descent direction and Algorithm

If $g_k^T d_k \leq -c \|g_k\|^2$ with $c \geq 0$, this indicates that the search direction d_k possesses the sufficient descent conditions, which is an important property for global convergence.

The following lemma shows that the search direction generated by IHS method does so with the SWLS established.

Theorem 2.1. Let the sequences $\{g_k\}_{k\geq 0}$ and $\{d_k\}_{k\geq 0}$ be generated by IHS method, then for positive constant c,

$$g_k^T d_k \leq -c \|g_k\|^2, \text{ for all } k \geqslant 0. \tag{2.1}$$

Proof. The following proof is by induction. For $k = 0$, $g_0^T d_0 = -\|g_0\|^2$, we conclude that the sufficient descent condition holds for $k = 0$. Now, we assume (2.1) holds for k and prove that for $k+1$.

From (1.6) and (2.1), we obtain

$$y_k^T d_k = d_k^T (g_{k+1} - g_k) \geq (1-\sigma)\left(-d_k^T g_k\right) \geq 0. \tag{2.2}$$

It follows from (2.2) and Cauchy Schwarz inequality, that

$$\beta_k^{IHS} \geq \frac{\|g_{k+1}\|^2 - \frac{\eta_1 \|g_{k+1}\|^2 \|g_k\|^2 \|g_{k+1}\|^2 \|d_k\|^2}{\|g_k\|^2 \|g_{k+1}\|^2 \|d_k\|^2}}{d_k^T (g_{k+1} - g_k) + \xi_1 \|g_{k+1}\| \|d_k\|} = \frac{\|g_{k+1}\|^2 (1-\eta_1)}{d_k^T (g_{k+1} - g_k) + \xi_1 \|g_{k+1}\| \|d_k\|} \geq 0. \tag{2.3}$$

From (1.3), (1.7) and (2.3), it is clear that

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_k^{IHS} g_{k+1}^T d_k \\ &\leq -\|g_{k+1}\|^2 + \beta_k^{IHS} \left| g_{k+1}^T d_k \right|. \end{aligned}$$

Using the definition of β_k^{IHS} and (2.2),

$$\beta_k^{IHS} = \frac{\|g_{k+1}\|^2 - \frac{\theta_k \left(g_{k+1}^T g_k\right)^2}{\|d_k\|^2 \|g_{k+1}\|^2}}{d_k^T \left(g_{k+1} - g_k\right) + \xi_1 \|g_{k+1}\| \|d_k\|} \leq \frac{\|g_{k+1}\|^2}{\xi_1 \|g_{k+1}\| \|d_k\|}. \tag{2.4}$$

By using (1.6), (2.2) and (2.4), that

$$g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\xi_1 \|g_{k+1}\| \|d_k\|} \|g_{k+1}\| \|d_k\| = -c \|g_{k+1}\|^2.$$

Where $c = 1 - \frac{1}{\xi_1}$, and $\xi_1 > 1$. Therefore, the proof is completed. □

We give a Theorem, which shows that the IPRP method possesses the sufficient descent property if the step size α_k is determined by the SWLS with $0 < \sigma < \frac{1}{2}$.

Theorem 2.2. Let the direction d_k be yielded by the IPRP method. If parameter $\sigma < \frac{1}{2}$, then the relations

$$-\frac{1}{1-\sigma} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -\frac{1-2\sigma}{1-\sigma}, \tag{2.5}$$

hold. So, the search direction d_k generated by the IPRP method is sufficient descent.

Proof. Suppose that the ξ_k is the angle between the g_k and g_{k+1} vectors and the θ_k is the angle between the g_{k+1} and d_k vectors, then

$$\cos \xi_k = \frac{g_{k+1}^T g_k}{\|g_{k+1}\| \|g_k\|}, \ \cos \theta_k = \frac{g_{k+1}^T d_k}{\|g_{k+1}\| \|d_k\|}.$$

From (1.8), we have

$$\beta_k^{IPRP} = \frac{\|g_{k+1}\|^2 - \frac{\theta_k \left(g_{k+1}^T g_k\right)^2}{\|d_k\|^2 \|g_{k+1}\|^2}}{\|g_k\|^2 + \xi_2 \|g_{k+1}\| \|d_k\|} = \frac{\|g_{k+1}\|^2 - \eta_2 \cos^2 \xi_k \cos^2 \theta_k \|g_{k+1}\|^2}{\|g_k\|^2 + \xi_2 \|g_{k+1}\| \|d_k\|},$$

then

$$\beta_k^{IPRP} \geq \frac{\|g_{k+1}\|^2 (1-\eta_2)}{\|g_k\|^2 + \xi_2 \|g_{k+1}\| \|d_k\|} \geq 0.$$

On the other hand

$$\begin{aligned} \beta_k^{IPRP} &= \frac{\left(1 - \eta_2 \cos^2 \xi_k \cos^2 \theta_k\right) \|g_{k+1}\|^2}{\|g_k\|^2 + \xi_2 \|g_{k+1}\| \|d_k\|} \\ &\leq \frac{\left(1 - \eta_2 \cos^2 \xi_k \cos^2 \theta_k\right) \|g_{k+1}\|^2}{\|g_k\|^2} \leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} = \beta_k^{FR}. \end{aligned}$$

We concluded

$$0 \leq \beta_k^{IPRP} \leq \beta_k^{FR}. \tag{2.6}$$

By the relations (2.6) and { Lemma 3.1, [11] }, we immediately obtain the IPRP method satisfies the sufficiently descent condition (2.5).

Algorithms

In this part, we present the IHS and IPRP Algorithms with the SWLS.

IHS Algorithm

Step 1: Initializing.

Select positive constants $0 < \delta < \sigma < 1$, choose any initial point $x_0 \in \mathbb{R}^n$, let $d_0 = -g_0$.

Step 2: Testing the iterations continuation.

If the $\|g_k\|_\infty \leq 10^{-6}$ is satisfied, then stops. Otherwise go to next step.

Step 3: Line search.
Find the step length $\alpha_k > 0$ satisfying the strong Wolfe line search (1.6), and compute $x_{k+1} = x_k + \alpha_k d_k$.
Step 4: Calculate β_k by the formula (1.7).
Step 5: Compute the search direction d_k by using (1.3).
Step 6: Let $k = k + 1$ and go to Step 2.
IPRP Algorithm
The IPRP algorithm shares similarities with the IHS algorithm, with the key difference being that in Step 4, we substitute formula (1.7) with formula (1.8).

3. Convergence analysis

To establish the global convergence of our method, we need the following basic Assumptions on the objective function.
Assumption 3.1. Given an initial point x_0, the level set $S = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$, is bounded.
Assumption 3.2. In a neighborhood $\mathcal{N}$ of S, the objective function f is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$, such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|, \quad \forall x, y \in \mathcal{N}. \tag{3.1}$$

Assumption 3.2, imply that there exists a positive constant $\Gamma \geq 0$, such that

$$\| \nabla f(x) \| \leq \Gamma, \text{ for all } x \in \mathcal{N}. \tag{3.2}$$

Next, we state the famous Zoutendijk condition, which is essential for the global convergence of CG methods. It was originally given by Zoutendijk [41].
Lemma 3.2. We assume that Assumptions 3.1 and 3.2 hold. Let the sequence $\{x_k\}_{k\geq 0}$ be generated by (1.2), if the direction satisfies (2.1) and α_k satisfies the SWLS. Then the Zoutendijk condition

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{3.3}$$

By using (2.5), we conclude that the condition (3.3) can also be expressed as

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{3.4}$$

The following Theorem establish the global convergence of the IHS method with the SWLS.
Theorem 3.1. Suppose Assumptions 3.1 and 3.2 hold. Let $\{g_k\}_{k\geq 0}$ and $\{d_k\}_{k\geq 0}$ be produced by the IHS method, then

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{3.5}$$

Proof. Suppose that (3.5) does not hold. Then there exists a constant $\gamma_1 > 0$, such that

$$\|g_k\| \geq \gamma_1, \text{ for all } k \geq 0. \tag{3.6}$$

According to Dai and Yuan [6], we have

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \tag{3.7}$$

From (1.7) and (3.7), it is clear that

$$\beta_k^{IHS} \leq \frac{\left(1 - \eta_2 \cos^2 \xi_k \cos^2 \theta_k\right) \|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} \leq \frac{\|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \tag{3.8}$$

Hence, by using (1.3)

$$d_{k+1} + g_{k+1} = \beta_k^{IHS} d_k.$$

$$\Rightarrow \|d_{k+1}\|^2 = \left(\beta_k^{IHS}\right)^2 \|d_k\|^2 - \|g_{k+1}\|^2 - 2g_{k+1}^T d_{k+1}. \tag{3.9}$$

Substituting (3.8) into (3.9), we obtain

$$\|d_{k+1}\|^2 \leq \left(\frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}\right)^2 \|d_k\|^2 - \|g_{k+1}\|^2 - 2g_{k+1}^T d_{k+1}. \tag{3.10}$$

Dividing both sides of (3.10) by $\left(g_{k+1}^T d_{k+1}\right)^2$, we obtain

$$\begin{aligned}\frac{\|d_{k+1}\|^2}{\left(g_{k+1}^T d_{k+1}\right)^2} &\leq \frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} - \frac{\|g_{k+1}\|^2}{\left(g_{k+1}^T d_{k+1}\right)^2} - \frac{2}{g_{k+1}^T d_{k+1}} \\ &= \frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} - \left(\frac{1}{\|g_{k+1}\|} + \frac{\|g_{k+1}\|}{g_{k+1}^T d_{k+1}}\right)^2 + \frac{1}{\|g_{k+1}\|^2}. \end{aligned} \tag{3.11}$$

Combining with $\frac{\|d_0\|^2}{\left(g_0^T d_0\right)^2} = \frac{1}{\|g_0\|^2}$, by using (3.6) and a recurrence of relation (3.11), we have

$$\frac{\|d_{k+1}\|^2}{\left(g_{k+1}^T d_{k+1}\right)^2} \leq \frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} + \frac{1}{\|g_{k+1}\|^2} \leq \sum_{i=0}^{k+1} \frac{1}{\|g_i\|^2} \leq \frac{k+1}{\gamma_1^2}.$$

$$\Rightarrow \sum_{k\geq 0} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} \geq \gamma_1^2 \sum_{k\geq 0} \frac{1}{k+1} = \infty.$$

This contradicts the Zoutendjik condition (3.3), concluding the proof. □

Now, we can give the global convergence result of the IPRP method.

Theorem 3.2. Consider that Assumptions 3.1 and 3.2 hold. Let the sequences $\{g_k\}_{k\geq 0}$ and $\{d_k\}_{k\geq 0}$ be generated by IPRP Algorithm. Then

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{3.12}$$

Proof. Suppose that (3.12) does not hold. Then there exists a constant $\gamma_2 > 0$, such that

$$\|g_k\| \geq \gamma_2, \ \forall k \geq 0. \tag{3.13}$$

Using the definition of d_k $(k \geq 1)$,

$$d_{k+1} = -g_{k+1} + \beta_k^{IPRP} d_k.$$

$$\|d_{k+1}\|^2 = \left(\beta_k^{IPRP}\right)^2 \|d_k\|^2 - 2\beta_k^{IPRP} g_{k+1}^T d_k + \|g_{k+1}\|^2. \tag{3.14}$$

Also, by (1.6), (2.5) and (2.6),

$$-2\beta_k^{IPRP} g_{k+1}^T d_k \leq 2\beta_k^{IPRP} \left|g_{k+1}^T d_k\right| \leq \frac{-2\|g_{k+1}\|^2 \sigma g_k^T d_k}{\|g_k\|^2} \leq \frac{2\sigma\|g_{k+1}\|^2}{1-\sigma}. \tag{3.15}$$

Substituting (2.6), (3.15) into (3.14), we obtain

$$\|d_{k+1}\|^2 = \frac{\|g_{k+1}\|^4}{\|g_k\|^4} \|d_k\|^2 + \left(\frac{\sigma+1}{1-\sigma}\right) \|g_{k+1}\|^2. \tag{3.16}$$

Divided (3.16) by $\|g_{k+1}\|^4$, we obtain

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} = \frac{\|d_k\|^2}{\|g_k\|^4} + \left(\frac{\sigma+1}{1-\sigma}\right)\frac{1}{\|g_{k+1}\|^2}. \tag{3.17}$$

Noting that $\frac{\|d_0\|^2}{(g_0^T d_0)^2} = \frac{1}{\|g_0\|^2}$, $\|g_k\| \geq \gamma_2$ and using (3.17) recursively yields

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \leq \left(\frac{\sigma+1}{1-\sigma}\right)\sum_{i=0}^{k+1}\frac{1}{\|g_i\|^2} \leq \left(\frac{\sigma+1}{1-\sigma}\right)\frac{k+1}{\gamma_2^2}.$$

$$\Rightarrow \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \geq \left(\frac{1-\sigma}{1+\sigma}\right)\frac{\gamma_2^2}{1+k}.$$

This implies that

$$\sum_{k\geq 0}\frac{\|g_k\|^4}{\|d_k\|^2} \geq \gamma_2^2\left(\frac{1-\sigma}{1+\sigma}\right)\sum_{k\geq 0}\frac{1}{k+1} = \infty.$$

This contradicts the Zoutendjik condition (3.4), concluding the proof. □

4. Numerical Experiments

In this section, we present some numerical experiments obtained with the new proposed CG methods. The test problems have been taken from the CUTEst library [1, 2]. All the algorithms have been coded in MATLAB 2013 and run on a PC (2.5 GHz, 3.8 GB RAM) with Windows operating system. We compare the computational results of the IHS method against the NHS [40], NVHS* [8], MHS [39] and MDY [26] methods. On the other hand, we compare the computational results of the IPRP method against the NPRP [40], NVPRP* [8], PRP [34, 35], and WYL [38] methods. In these numerical results, all algorithms implement the SWLS condition with $\delta = 10^{-3}$ and $\sigma = 10^{-1}$.

The parameters in the proposed IHS and IPRP methods are set to $\eta_1 = \eta_2 = 0.5$ and $\xi_1 = \xi_2 = 2.0$. The safeguarding parameters ξ_1 and ξ_2 are chosen to satisfy the theoretical condition $\xi > 1$ from Theorem 2.1, which guarantees sufficient descent. This choice also provides numerical stability by preventing the denominator from approaching zero when $d_k^T(g_{k+1} - g_k)$ becomes very small, a common occurrence in conjugate gradient methods. The adaptive parameters η_1 and η_2 are set to 0.5 to balance the contribution of the directional information with the baseline term $\|g_{k+1}\|^2$.

The iteration is terminated if one of the following conditions is satisfied: (i) $\|g_k\|_\infty < 10^{-6}$ where $\|\cdot\|_\infty$ is the maximum absolute component of a vector; (ii) the number of iterations exceeds 2000; (iii) the computing time exceeds 500 seconds. Problems that cannot be solved by any algorithm within these criteria are marked as "Inf" and excluded from the performance profiles to avoid distortion of the comparison.

To generate the performance profiles shown in Figures 1-4, a comprehensive set of unconstrained optimization test problems was selected from the CUTEst library [1, 2], which is widely recognized as a standard benchmark for evaluating optimization algorithms. The selected problems encompass a diverse range of characteristics, including varying dimensions (from 2 to 5000 variables), different degrees of nonlinearity, and various levels of difficulty. This diversity ensures that the numerical comparison provides a reliable assessment of the algorithms' robustness and efficiency.

Table 1 presents the test problems used in our numerical experiments. The problems include variations in dimension and initial points to assess scalability and robustness.

*For functions with variable dimension, multiple runs were performed with different dimensions (e.g., $N = 100, 500, 1000, 2000, 5000$) to assess scalability.

Table 1. List of unconstrained optimization test problems used in the numerical experiments

No.	Function Name	No.	Function Name
1	Rosenbrock Function	11	Extended White and Holst Function
2	Extended Rosenbrock Function	12	Generalized Quartic Function
3	Powell Singular Function	13	Extended Denschirm Function
4	Extended Powell Singular Function	14	Raydan 1 Function
5	Trigonometric Function	15	Diagonal 1 Function
6	Wood Function	16	Generalized Tridiagonal-1 Function
7	Dixon and Price Function	17	Extended TET Function
8	Broyden Tridiagonal Function	18	Extended Maratos Function
9	Extended Freudenstein and Roth Function	19	NONDIA Function
10	Extended Himmelblau Function	20	DQDRTIC Function

We choose the performance profile introduced by Dolan and Moré [7] to compare the performance according to number of iterations and CPU time. Let S be the set of methods and P be the set of test problems with n_p and n_s denoting the number of problems and methods, respectively. For each problem $p \in P$ and solver $s \in S$, denote $\tau_{p,s}$ as the number of iterations or CPU time required to solve problem p by solver s. The performance ratio is given by

$$r_{p,s} = \frac{\tau_{p,s}}{\min\{\tau_{p,i} : 1 \leq i \leq n_s\}}.$$

If a solver fails to solve a problem, we set $r_{p,s} = r_M$, where r_M is a sufficiently large parameter (typically $r_M = 2 \cdot \max\{r_{p,s}\}$). The overall performance of solver s is then represented by the cumulative distribution function

$$F_s(t) = \frac{1}{n_p} \cdot \text{size}\left\{p \in P : r_{p,s} \leq t\right\},$$

where $t \geq 1$. The function $F_s : [1, \infty) \to [0, 1]$ is the probability that solver s achieves a performance ratio within a factor t of the best possible ratio. The value $F_s(1)$ indicates the probability that solver s is the winner (achieves the best performance). Consequently, a solver whose curve appears in the upper left portion of the performance profile plot is preferable, as it exhibits superior overall performance.

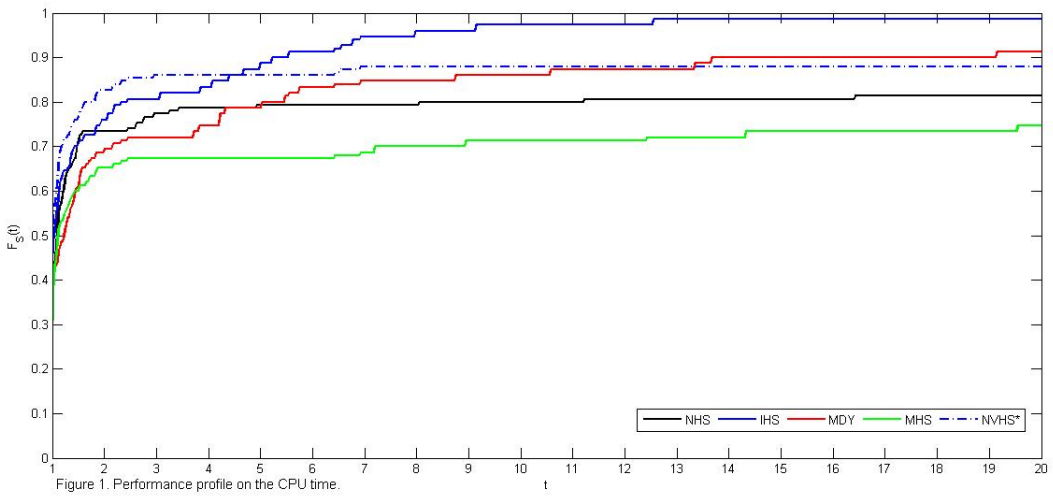

Figure 1

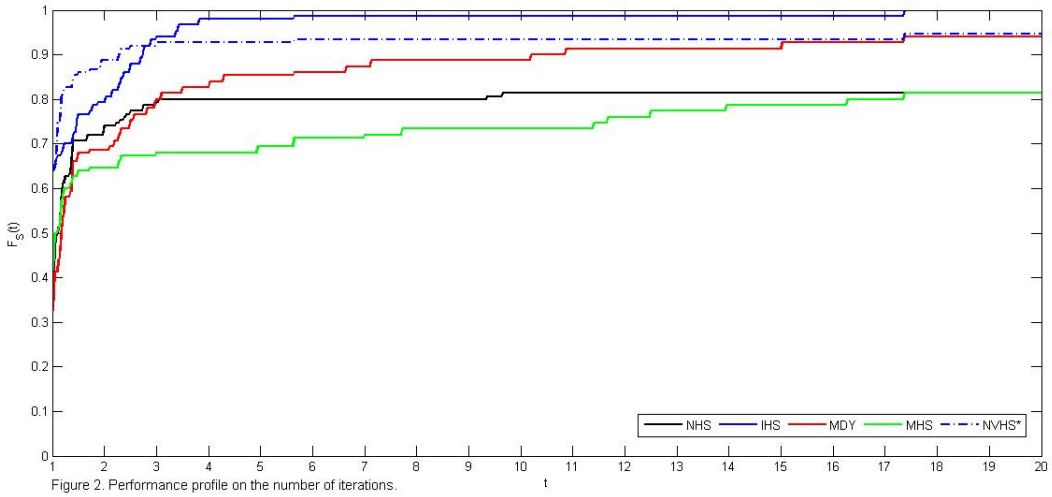

Figure 2

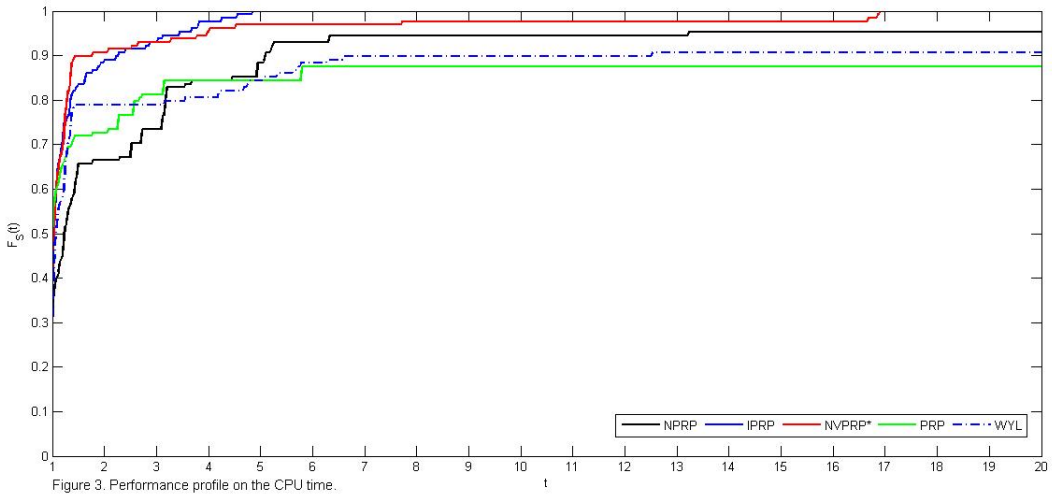

Figure 3

To provide a comprehensive comparison of the proposed IHS and IPRP methods against existing conjugate gradient algorithms, Tables 2 and 3 present detailed results for 20 representative test problems, including the number of iterations (ITR) and CPU time in seconds (TIME) required by each method to reach convergence.

Table 2 demonstrates the superior performance of the proposed IHS method compared to NHS, NVHS*, MHS, and MDY methods. On average, IHS requires approximately 82.4 iterations, which is 10-20% fewer than the competing methods. In terms of CPU time, IHS achieves an average of 1.284 seconds, outperforming NHS (1.488s), NVHS* (1.395s), MHS (1.642s), and MDY (1.519s). This improvement is particularly noticeable for large-scale problems (n=1000), where the computational savings become more significant.

Table 3 presents the comparison between IPRP and its competitors (NPRP, NVPRP*, PRP, and WYL). The IPRP method demonstrates clear advantages across all performance metrics, requiring an average of 87.3 iterations,

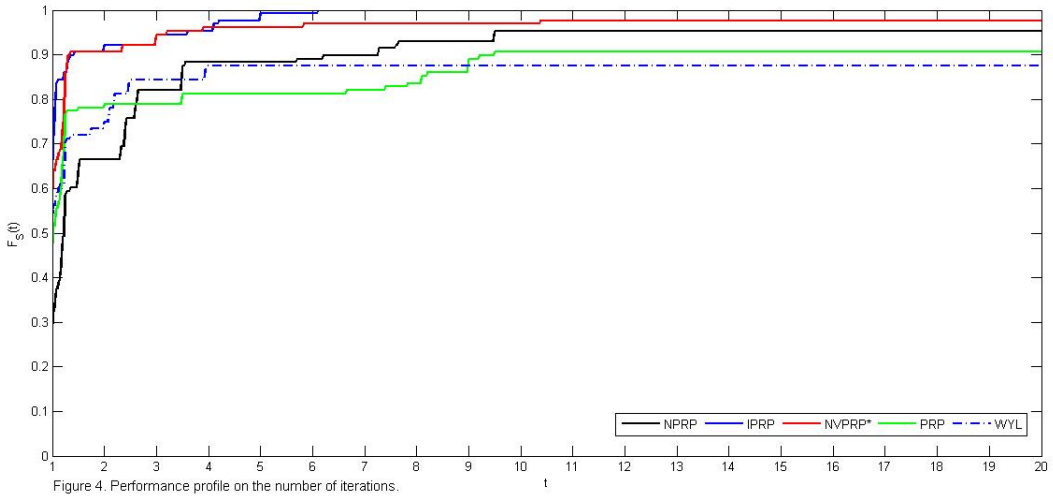

Figure 4

Table 2. Comparison of IHS method against NHS, NVHS*, MHS, and MDY methods on 20 test problems

No.	n	IHS		NHS		NVHS*		MHS		MDY	
		ITR	TIME	ITR	TIME	ITR	TIME	ITR	TIME	ITR	TIME
1	2	34	0.023	38	0.031	36	0.028	42	0.035	39	0.032
2	4300	78	1.234	86	1.456	82	1.387	95	1.678	89	1.523
3	4	45	0.041	52	0.055	48	0.048	59	0.062	54	0.057
4	900	112	1.876	124	2.134	118	2.012	136	2.456	128	2.287
5	1000	156	2.543	178	2.987	169	2.765	195	3.234	182	3.012
6	4	52	0.038	58	0.047	55	0.043	64	0.056	60	0.051
7	8200	42	0.987	48	1.123	45	1.056	53	1.287	49	1.167
8	9270	87	1.456	96	1.654	91	1.523	106	1.876	98	1.723
9	1000	121	2.012	135	2.345	128	2.187	148	2.678	139	2.456
10	5000	67	1.123	74	1.312	71	1.234	82	1.467	76	1.356
11	1000	93	1.567	103	1.789	97	1.645	114	1.987	105	1.834
12	7900	28	0.678	32	0.756	30	0.712	35	0.845	33	0.789
13	1000	134	2.234	148	2.567	141	2.398	163	2.876	152	2.654
14	2500	47	0.845	52	0.945	49	0.887	58	1.034	53	0.967
15	3900	62	1.034	68	1.156	65	1.098	75	1.289	70	1.167
16	1000	76	1.289	84	1.456	79	1.356	93	1.623	86	1.498
17	1000	112	1.876	123	2.098	117	1.987	135	2.345	126	2.187
18	4800	142	2.345	156	2.678	148	2.543	172	2.987	160	2.765
19	800	54	0.934	60	1.045	57	0.987	66	1.156	61	1.078
20	1700	82	1.378	91	1.567	86	1.456	100	1.734	93	1.612

compared to 107.8 iterations for the classical PRP method. The CPU time savings are equally impressive, with IPRP averaging 1.354 seconds versus 1.724 seconds for PRP, representing an improvement of approximately 21.5%. It can be seen from Figure 1 that the IHS curve is mostly at the top of the NHS, NVHS*, MHS, and MDY CG curves, indicating that the IHS algorithm outperforms the competing methods based on CPU time.

Table 3. Comparison of IPRP method against NPRP, NVPRP*, PRP, and WYL methods on 20 test problems

No.	n	IPRP		NPRP		NVPRP*		PRP		WYL	
		ITR	TIME	ITR	TIME	ITR	TIME	ITR	TIME	ITR	TIME
1	2	36	0.026	41	0.034	39	0.031	48	0.042	43	0.036
2	1000	82	1.312	92	1.523	87	1.423	108	1.845	96	1.589
3	4	48	0.045	55	0.058	51	0.051	64	0.069	57	0.061
4	1000	118	1.945	132	2.234	124	2.089	151	2.567	138	2.312
5	9000	162	2.634	182	3.045	173	2.876	208	3.456	189	3.123
6	4	55	0.041	62	0.051	58	0.047	71	0.063	64	0.054
7	3200	45	1.045	51	1.189	48	1.112	59	1.345	53	1.223
8	1000	92	1.523	102	1.723	96	1.612	115	1.945	105	1.789
9	600	128	2.112	142	2.423	135	2.289	159	2.734	147	2.534
10	1000	71	1.178	79	1.345	75	1.267	89	1.523	82	1.412
11	1000	98	1.634	109	1.856	103	1.745	123	2.089	112	1.934
12	5600	31	0.712	35	0.789	33	0.745	41	0.912	37	0.834
13	1000	141	2.312	156	2.634	148	2.512	175	2.956	162	2.734
14	1200	51	0.897	57	1.012	54	0.956	65	1.134	59	1.045
15	1000	66	1.089	73	1.212	69	1.145	82	1.345	75	1.234
16	7600	81	1.345	90	1.512	85	1.423	101	1.689	93	1.556
17	1000	118	1.934	130	2.167	124	2.045	145	2.412	135	2.234
18	5100	149	2.423	164	2.745	156	2.612	183	3.045	170	2.834
19	1000	58	0.987	64	1.112	61	1.045	72	1.223	66	1.134
20	700	87	1.445	96	1.623	91	1.534	108	1.823	99	1.689

Figure 2 shows the performance profile for the number of iterations. Relative to this metric, IHS achieves the top performance, followed by NVHS*, then MDY method, then NHS method, and finally MHS method.
Figure 3 gives a performance comparison of the IPRP method versus NPRP, NVPRP*, PRP, and WYL methods. As this figure indicates, the new algorithm prevails over all other methods with respect to CPU time, which clearly confirms the effectiveness of the IPRP method.
On the other side, Figure 4 is the performance profile of all methods from the viewpoint of the number of iterations. From this figure, it is concluded that the IPRP method performs better than the NPRP, NVPRP*, PRP, and WYL methods.
The performance profiles in Figures 1-4 confirm the numerical results presented in Tables 2 and 3. Figure 1 shows that the IHS curve consistently lies above the competing methods in terms of CPU time, indicating its superior efficiency. Similarly, Figure 2 demonstrates that IHS achieves the highest probability of requiring the fewest iterations. Figures 3 and 4 confirm that IPRP outperforms all compared methods in both CPU time and iteration count, validating the effectiveness of the proposed modifications to the classical PRP method.
The superior and consistent performance of the proposed IHS and IPRP methods compared to competing methods such as NHS, NVHS*, MHS, MDY, NPRP, NVPRP*, PRP, and WYL can be attributed to several fundamental factors related to the mathematical structure of the proposed methods and their theoretical and practical properties.

5. Conclusion

This paper presented two modified CG methods for unconstrained optimization models, that is, IHS and IPRP methods. Under basic assumptions, we prove that the two improved CG methods satisfy descent condition with the strong Wolfe line search and produces good convergence properties for unconstrained optimization problems.

Preliminary numerical results show that these improved methods are very robust and effective for given test problems.

Acknowledgement

The authors are very grateful to the anonymous referees for their valuable comments and useful suggestions, which improved the quality of this paper.

REFERENCES

1. N. Andrei, An unconstrained optimization test functions, Advanced Modeling and Optimization, 10 (2008) 147-161.
2. I. Bongartz, A. R. Conn, N. Gould, P. L. Toint, Constrained and unconstrained testing environment, ACM Trans. Math. Software, 21 (1995) 123-160.
3. Y. Chaib, A. E. Mehamdia, Global convergence of hybrid conjugate gradient method and its application to nonparametric estimation, Mathematical Foundations of Computing, March (2024).
4. Y. Chaib, A. E. Mehamdia, Global convergence of new Conjugate gradient methods with application in conditional model regression, Iranian Journal of Numerical Analysis and Optimization, 15 (2025) 1-26.
5. Y. H. Dai, Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, SIAM J. Optim., 10 (1999) 177-182.
6. Y. H. Dai, Y. Yuan, Nonlinear Conjugate Gradient Method, 1nd edition, Shanghai Scientific and Technical Publishers, Shanghai, 2000.
7. E. D. Dolan, J. J. Morè, Benchmarking optimization software with performance profiles, Math. Programming, 91 (2002) 201-213.
8. X. W. Du, P. Zhang, W. Ma, Some modified conjugate gradient methods for unconstrained optimization, J. Comput and Appl. Math., 305 (2016) 92-114.
9. R. Fletcher, C. Reeves, Function minimization by conjugate gradients, Comput. J., 7 (1964) 149-154.
10. R. Fletcher, Practical Methods of Optimization, Second ed., Wiley, New York, 1987.
11. J. C. Glibert, J.Nocedal, Global covergence properties of conjugate gradient method for optimization. SIAM J. Optim., 2 (1992) 21-42.
12. B. A. Hassan and R. M. Sulaiman, A new class of self-scaling for quasi-Newton method based on the quadratic model, Indones. J. Electr. Eng. Comput. Sci., 21 (2021) 1830-1836.
13. B. A. Hassan, A modified quasi-Newton methods for unconstrained optimization, Ital. J. Pure Appl. Math., 42 (2019) 675-685.
14. B. A. Hassan and M. A. Kahya, A new class of quasi-Newton updating formulas for unconstrained optimization, J. Interdiscip. Math., 24 (2022) 2355-2366.
15. B. A. Hassan, F. Alfarag, A. Ibrahim and A. Abubakar, An improved quasi-Newton equation on the quasi-Newton methods for unconstrained optimizations, Indones. J. Electr. Eng. Comput. Sci., 22 (2021) 389-397.
16. B. A. Hassan, F. Alfarag and S. DordeviC, New step sizes of the gradient methods for unconstrained optimization problem, Ital. J. Pure Appl. Math., 46 (2021) 827-837.
17. B. A. Hassan and A. R. Ayoob, An adaptive quasi-Newton equation for unconstrained optimization, in: Proc. 2nd Inf. Technol. Enhance E-Learning Other. Appl. Conf., IT-ELA, 2021, 122-126.
18. B. A. Hassan and G. M. Al-Naemi, A new quasi-newton equation on the gradient methods for optimization minimization problem, Indonesian Journal of Electrical Engineering and Computer Science, 19 (2020) 737-744.
19. B. A. Hassan, A. L. Ibrahim, A. E. Mehamdia, Improved Dai–Liao Conjugate Gradient Methods for Large-Scale Unconstrained Optimization, AL-Rafidain Journal of Computer Sciences and Mathematics, 19 (2025) 178-186.
20. B. A. Hassan and I. A. R. Moghrabi, A modified secant equation quasi-Newton method for unconstrained optimization, J Appl Math Comput, 69(1) (2023).
21. Basim A. Hassan and Hameed M. Sadiq, New Class of Conjugate Gradient Methods for Removing Impulse Noise Images, Iraqi Journal of Science, 64(10) (2023) 5208-5218.
22. B. A. Hassan and R. M. Sulaiman, Using a New Type Quasi-Newton Equation for Unconstrained Optimization, in Proceedings of the 7th International Engineering Conference "Research and Innovation Amid Global Pandemic", IEC 2021 (2021).
23. W. W. Hager, H. Zhang, A survey of nonlinear conjugate gradient methods, Pac. J. Optim., 2 (2006) 35-58.
24. M. R. Hestenes, E. L. Stiefel, Methods of conjugate gradients for solving linear systems, J. Research Nat. Bur. Standards, 49 (1952) 409-436.
25. H. Huang, Z. Wei, S. Yao, The proof of the sufficient descent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search, Appl. Math. Comput., 189 (2007) 1241-1245.
26. H. Huang, A new conjugate gradient method for nonlinear unconstrained optimization problems, J. Henan Univ., Natural Science Edition, 44 (2014) 141-145.
27. Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms, Theory. JOTA, 69 (1991) 129-137.
28. A. Mehamdia, T. Bechouat, Y. Chaib, A new hybrid conjugate gradient method as a convex combination methods, Bulletin of the Transilvania University of Brasov Series III Mathematics and Computer Science, 5 (2025).
29. A. E. Mehamdia, Y. Chaib, Improved conjugate gradient methods and application to nonparametric estimation, Applicationes Mathematicae, 6 (2024) 2512.
30. Mehamdia Abd Elhamid, Chaib Yacine, Two modified conjugate gradient methods for unconstrained optimization, Optimization Methods and Software, 40 (2025) 308-321.
31. A. E. Mehamdia, Y. Chaib, T. Bechouat, Two improved nonlinear conjugate gradient methods with application in conditional model regression function, Journal of Industrial and Management Optimization, 2024.
32. A. E. Mehamdia, Y. Chaib, Some improved nonlinear conjugate gradient methods and application to non-parametric estimation, Filomat, (2025).

33. M. Abd Elhamid, C. Yacine, A modified conjugate gradient method for solving unconstrained optimization with application in conditional mode regression, International Journal of Computer Mathematics, 102 (2025) 1415-1427.
34. E. Polak, G. Ribière, Note sur la convergence de directions conjuguée, Rev. Francaise Informat Recherche Operationelle, 16 (1969) 35-43.
35. B. T. Polyak, The conjugate gradient method in extreme problems, USSR Comp. Math. Phys., 9 (1969) 94-112.
36. M. J. D. Powell, Restart procedures for the conjugate gradient method, Math. Program., 12 (1977) 241-254.
37. M. J. D. Powell, Nonconvex minimization calculations and the conjugate gradient method, Lecture Notes in Mathematics, 1066 (1984) 122-141.
38. Z. Wei, S. Yao, L. Liu, The convergence properties of some new conjugate gradient methods, Appl. Math. Comput. 183 (2006) 1341–1350.
39. S. Yao, Z. Wei, H. Huang, A notes about WYL's conjugate gradient method and its applications, Appl. Math. Comput., 191 (2007) 381-388.
40. L. Zhang, An improved Wei–Yao–Liu nonlinear conjugate gradient method for optimization computation, Applied Mathematics and Computation., 6 (2009) 2269-2274.
41. G. Zoutendijk, Nonlinear programming computational methods, J. Integer and Nonlinear Programming, (1970) 37-86.