RAIRO-Oper. Res. 58 (2024) 613–627
https://doi.org/10.1051/ro/2023196

RAIRO Operations Research
www.rairo-ro.org

A NEW FAMILY OF HYBRID CONJUGATE GRADIENT METHOD FOR UNCONSTRAINED OPTIMIZATION AND ITS APPLICATION TO REGRESSION ANALYSIS

Sabrina Ben Hanachi*, Badreddine Sellami and Mohammed Belloufi

Abstract. We know many conjugate gradient algorithms (CG) for solving unconstrained optimization problems. In this paper, based on the three famous Liu–Storey (LS), Fletcher–Reeves (FR) and Polak–Ribiére–Polyak (PRP) conjugate gradient methods, a new hybrid CG method is proposed. Furthermore, the search direction satisfies the sufficient descent condition independent of the line search. Likewise, we prove, under the strong Wolfe line search, the global convergence of the new method. In this respect, numerical experiments are performed and reported, which show that the proposed method is efficient and promising. In virtue of this, the application of the proposed method for solving regression models of COVID-19 is provided.

Mathematics Subject Classification. 65K05, 90C25, 90C26, 90C27, 90C30.

Received February 5, 2023. Accepted December 20, 2023.

1. Introduction

Unconstrained optimization is a branch of optimization in which we minimize an objective function that depends on real variables with the total absence of restrictions on their values of those variables. Thus, we consider the general unconstrained optimization problems, as follows:

$$\min\{f(x), x \in \mathbb{R}^n\}, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is the continuously differentiable function and its first derivative is represented by $g(x) = \nabla f(x)$. Though several robust optimization algorithms with rapid convergence have shown to be available to solve the above nonlinear optimization model, many researchers still refer to the conjugate gradient algorithm (CG) as it uses low memory and good convergence properties. Besides, this method was first established by Hestenes and Stiefel [18] to solve unconstrained linear optimization problems. Then, in 1964, Fletcher and Reeves [14] extended the form of the conjugate gradient method to solve unconstrained nonlinear minimization problems. As a consequence, the results of the expansion inspired researchers to suggest a new conjugate gradient method with good computational performance and, at the same time, good convergence properties [6]. Generally,

Keywords. Unconstrained optimization, hybrid conjugate gradient method, sufficient descent, convex combination, global convergence.

Laboratory Informatics and Mathematics (LiM), Mohamed Cherif Messaadia University, Souk-Ahras, Algeria.
*Corresponding author: s.benhanachi@univ-soukahras.dz; sabrinabenhannachi9@gmail.com

the iterates of the CG methods are usually determined through the following recursive computational scheme:

$$x_{k+1} = x_k + \alpha_k d_k. \quad k = 0, 1, \dots n, \tag{2}$$

and

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

where x_k is the current iteration, g_k is the gradient of f at the point x_k, d_k is the search direction, $\beta_k \in \mathbb{R}$ is the conjugate parameter which characterizes different versions of the CG methods and $\alpha_k > 0$ is the step size many line search techniques can obtain. In the midst of which, exact line search, weak Wolfe line search, or strong Wolfe line search, but we use, in our research, strong Wolfe line search, which is defined by the following conditions [27, 28]:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{4}$$

$$\left| g_{k+1}^T d_k \right| \leq -\sigma g_k^T d_k, \tag{5}$$

where scalars δ and σ satisfy $0 < \delta \leq \sigma < 1$.

As we said, the CG algorithms differ by choosing the coefficient β_k. The most common standard CG methods are the Hestenes–Stiefel (HS) method [18], the Fletcher–Reeves (FR) method [14], the Conjugate-Descent (CD) method [13], the Dai–Yuan (DY) method [8], the Liu–Storey (LS) method [22] and the Polak–Ribiére–Polyak (PRP) method [23, 24], respectively, which are defined as:

$$\beta_k^{\text{HS}} = \frac{g_k^T y_{k-1}}{y_{k-1}^T d_{k-1}} \qquad \beta_k^{\text{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \qquad \beta_k^{\text{CD}} = \frac{\|g_k\|^2}{-g_{k-1}^T d_{k-1}}$$

$$\beta_k^{\text{DY}} = \frac{\|g_k\|^2}{y_{k-1}^T d_{k-1}} \qquad \beta_k^{\text{LS}} = \frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}} \qquad \beta_k^{\text{PRP}} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2},$$

where $y_{k-1} = g_k - g_{k-1}$ and $\|\cdot\|$ stands for the Euclidean norm. If f is a strictly convex quadratic function, all the above methods are equivalent, while they behave differently for general non-quadratic functions.

One of the essential classes of CG methods is the hybrid conjugate gradient algorithms. Further, the hybrid schemes have better computational performances and more robust convergence properties than conventional CG methods as they take advantage of the two parameters used to build them. For this reason, many researchers cared about hybrid or mixed conjugate gradient methods. Djordjević [10], proposed the following hybrid method: $\beta_k^{\text{hyb}} = \theta_k \beta_k^{\text{FR}} + (1 - \theta_k)\beta_k^{\text{HS}}$, Andrei [5], proposed the following hybrid method: $\beta_k^c = \theta_k \beta_k^{\text{DY}} + (1 - \theta_k)\beta_k^{\text{HS}}$, Li and Sun [20], proposed the following hybrid method: $\beta_k^N = \theta_k \beta_k^{\text{MMWU}} + (1 - \theta_k)\beta_k^{\text{FR}}$, Liu and Li [21], proposed the following hybrid method: $\beta_k^{\text{hyb}} = \theta_k \beta_k^{\text{DY}} + (1 - \theta_k)\beta_k^{\text{LS}}$. In addition, Sabrina *et al.* [17] proposed a new hybrid CG method based on combination of FR, PRP and DY conjugate gradient algorithms in which

$$\beta_k^{\text{hyb}} = \delta_k \beta_k^{\text{FR}} + \gamma_k \beta_k^{\text{PRP}} + (1 - \delta_k - \gamma_k)\beta_k^{\text{DY}},$$

where

$$\gamma_k = -\frac{g_k^T g_{k+1} \|g_k\|^2 + \delta_k \left(y_k^T s_k - \|g_k\|^2\right) \|g_{k+1}\|^2}{\left(g_{k+1}^T y_k\right)\left(y_k^T s_k\right) - \|g_{k+1}\|^2 \|g_k\|^2}, \quad \text{where, } 0 < \delta_k < 1.$$

Inspired by this research, we propose in this study a new hybrid CG method based on combination of LS, FR and PRP conjugate gradient algorithms for solving unconstrained optimization problems. In addition, we alike apply, in this study, the new method for solving a model of COVID-19 outbreak around the globe in which the data is taken from January to September 2020.

In light of this, the paper is structured as follows. In Section 2, we will describe the proposed method with its corresponding algorithm and further establish the descent condition and convergence under inexact line search. In Section 3, we present the numerical experiments to show the efficiency of our new method and the application of regression models of COVID-19 using the new method is illustrated in Section 4. Finally, a brief conclusion is drawn in Section 5.

2. Proposed method, algorithm

In this paper, we propose another combination of LS, FR and PRP conjugate gradient algorithms. We use the following conjugate gradient parameter:

$$\beta_k^{\text{New}} = \delta_k \beta_k^{\text{LS}} + \gamma_k \beta_k^{\text{FR}} + (1 - \delta_k - \gamma_k)\beta_k^{\text{PRP}}. \tag{6}$$

As a consequence, the direction d_k is given by:

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k^{\text{New}} d_{k-1}, & \text{if } k \geq 1. \end{cases} \tag{7}$$

The parameters δ_k, γ_k in (6) satisfying $0 \leq \delta_k, \gamma_k \leq 1$ which will be determined in a particular way that will later be described. It should be noted that:

- If $\delta_k = 1$ and $\gamma_k = 0$, then $\beta_k^{\text{New}} = \beta_k^{\text{LS}}$.
- If $\delta_k = 0$ and $\gamma_k = 1$, then $\beta_k^{\text{New}} = \beta_k^{\text{FR}}$.
- If $\delta_k = 0$ and $\gamma_k = 0$, then $\beta_k^{\text{New}} = \beta_k^{\text{PRP}}$.
- If $\delta_k = 0$ and $0 < \gamma_k < 1$, then $\beta_k^{\text{New}} = \gamma_k \beta_k^{\text{FR}} + (1 - \gamma_k)\beta_k^{\text{PRP}}$ *i.e.* β_k^{New} is a convex combination of β_k^{FR} and β_k^{PRP}. See [9].
- If $\gamma_k = 0$ and $0 < \delta_k < 1$, then $\beta_k^{\text{New}} = \delta_k \beta_k^{\text{LS}} + (1 - \delta_k)\beta_k^{\text{PRP}}$ *i.e.* β_k^{New} is a convex combination between β_K^{LS} and β_K^{PRP}. See [1].
- If $1 - \delta_k - \gamma_k = 0$, $0 < \delta_k, \gamma_k < 1$, then $\gamma_k = 1 - \delta_k$. Then $\beta_k^{\text{New}} = \delta_k \beta_k^{\text{LS}} + (1 - \delta_k)\beta_k^{\text{FR}}$ *i.e.* β_K^{New} is a convex combination between β_k^{LS} and β_k^{FR}. See [11].

Finally, if $\delta_k \in]0, 1[$, $\gamma_k \in]0, 1[$ and $0 < \delta_k + \gamma_k < 1$, then we have a *new hybrid CG method* as a convex combination of three methods "LS, FR and PRP". From (6) and (7), it is clear that:

$$d_k = \begin{cases} -g_k, & k = 0 \\ -g_k + \delta_k \frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}} d_{k-1} + \gamma_k \frac{\|g_k\|^2}{\|g_{k-1}\|^2} d_{k-1} + (1 - \delta_k - \gamma_k) \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} d_{k-1}, & k \geq 1. \end{cases} \tag{8}$$

To select the parameters δ_k and γ_k, we use the traditional conjugacy condition, *i.e.* ($d_{k+1}^T y_k = 0$). Thus, we have the following lemma.

Lemma 1. *If the conjugacy condition $d_{k+1}^T y_k = 0$ is satisfied at every iteration, we get*

$$\delta_k = \frac{\left(g_{k+1}^T y_k (\|g_k\|^2 - d_k^T y_k) - \gamma_k (g_{k+1}^T g_k)(d_k^T y_k)\right)}{(\|g_k\|^2 + d_k^T g_k)(-g_{k+1}^T y_k)(d_k^T y_k)} d_k^T g_k, \qquad \textit{where } \; 0 < \gamma_k < 1. \tag{9}$$

Proof. Multiplying (8) by y_k from the left and using the conjugacy condition, we obtain

$$0 = -g_{k+1}^T y_k + \delta_k \beta_k^{\text{LS}} d_k^T y_k + \gamma_k \beta_k^{\text{FR}} d_k^T y_k + (1 - \delta_k - \gamma_k)\beta_k^{\text{PRP}} d_k^T y_k,$$

$$g_{k+1}^T y_k = \delta_k \left(\frac{g_{k+1}^T y_k}{-g_k^T d_k}\right) d_k^T y_k + \gamma_k \left(\frac{\|g_{k+1}\|^2}{\|g_k\|^2}\right) d_k^T y_k + (1 - \delta_k - \gamma_k) \left(\frac{g_{k+1}^T y_k}{\|g_k\|^2}\right) d_k^T y_k.$$

Finally, after some algebra, we have:

$$\delta_k = \frac{\left(g_{k+1}^T y_k (\|g_k\|^2 - d_k^T y_k) - \gamma_k (g_{k+1}^T g_k)(d_k^T y_k)\right)}{(\|g_k\|^2 + d_k^T g_k)(-g_{k+1}^T y_k)(d_k^T y_k)} d_k^T g_k, \qquad \text{where } \; 0 < \gamma_k < 1.$$

□

The parameter δ_k given by (9) can be outside the interval $[0,1]$. However, in order to have a real convex combination in (6), the following rule is considered: if $\delta_k \leq 0$ then set $\delta_k = 0$ in (6), if $\delta_k \geq 1$, then take $\delta_k = 1$ in (6), if $\delta_k + \gamma_k \geq 1$, put $\delta_k + \gamma_k = 1$ in (6). Therefore, under this rule for δ_k selection.

Next, we give the algorithm of our proposed method below:

Algorithm (SCH)

Step 1: Initialization**.** Given $x_0 \in \mathbb{R}^n$ and the parameters $0 < \delta \leq \sigma < 1$. Set $k = 0$. Compute $f(x_0)$, $g_0 = \nabla f(x_0)$. Consider $d_0 = -g_0$, set the initial guess: $\alpha_0 = 0$ and $\gamma_k = 0.5$.
Step 2: Test a criterion for stopping iterations**.** If $\|g_k\| < 10^{-6}$, then stop. Else continue with Step 3.
Step 3: Line search**.** Compute $\alpha_k > 0$ $(k \neq 0)$ by the strong Wolfe line search, *i.e.*, α_k satisfies (4), (5).
Step 4: Generate**.** $x_{k+1} = x_k + \alpha_k d_k$. Compute $f(x_{k+1})$, $g_{k+1} = \nabla f(x_{k+1})$ and $y_k = g_{k+1} - g_k$.
Step 5. Compute δ_k as in equation (9).
Step 6. Calculate β_k^{New} by equation (6).
Step 7: Computation of the search direction**.** Compute $d = -g_{k+1} + \beta_k^{\text{New}} d_k$. If the restart criterion of Powell

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

is satisfied, then $d_{k+1} = -g_{k+1}$. Otherwise define $d_{k+1} = d$.
Step 8. Put $k = k+1$ and continue with Step 2.

2.1. The sufficient descent condition

In this study, we will establish the sufficient descent of our new method, which plays a vital role in the global convergence analysis. Thus, we need the following assumptions:

Assumption 1. *The level set $S = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$ is bounded, i.e. there exists a constant $B > 0$, such that*

$$\|x\| \leq B, \text{ for all } x \in S. \tag{10}$$

Assumption 2. *In a neighborhood $\mathcal{N}$ of S the function f is continuously differentiable and its gradient $\nabla f(x)$ is Lipschitz continuous, i.e. there exists a constant $0 < L < \infty$, such that:*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \text{for all } x, y \in \mathcal{N}. \tag{11}$$

Under Assumptions 1 and 2 on f, there exists a constant $\Gamma \geq 0$, such that:

$$\|\nabla f(x)\| \leq \Gamma, \tag{12}$$

for all $x \in S$ [4].

The following theorem proves that the search direction obtained by the new method satisfies the sufficient descent condition.

Theorem 1. *Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by SCH method. Then, the search direction d_k satisfies the sufficient descent condition:*

$$g_k^T d_k \leq -K\ \|g_k\|^2, \quad \forall k \geq 0, \tag{13}$$

where $K > 0$.

Proof. The following proof is by induction. We show that search direction d_k shall satisfy the sufficient descent condition holds for $k = 0$, the proof is a trivial one, *i.e.* $d_0 = g_0$ so $g_0^T d_0 = -\|g_0\|^2$, and we conclude that sufficient descent condition holds for $k = 0$. Next, we assume that (13) holds for some $k \geq 1$.

Now we have:

$$\begin{aligned} d_{k+1} &= -g_{k+1} + \beta_k^{\text{New}} d_k, \\ &= -g_{k+1} + \big(\delta_k \beta_k^{\text{LS}} + \gamma_k \beta_k^{\text{FR}} + (1-\delta_k-\gamma_k)\beta_k^{\text{PRP}}\big) d_k, \end{aligned}$$

we can write

$$d_{k+1} = -(\delta_k g_{k+1} + \gamma_k g_{k+1} + (1-\delta_k-\gamma_k) g_{k+1}) + \big(\delta_k \beta_k^{\text{LS}} + \gamma_k \beta_k^{\text{FR}} + (1-\delta_k-\gamma_k)\beta_k^{\text{PRP}}\big) d_k.$$

It follows that

$$d_{k+1} = \delta_k\big(-g_{k+1} + \beta_k^{\text{LS}} d_k\big) + \gamma_k\big(-g_{k+1} + \beta_k^{\text{FR}} d_k\big) + (1-\delta_k-\gamma_k)\big(-g_{k+1} + \beta_k^{\text{PRP}} d_k\big).$$

Produces after some arrangements

$$d_{k+1} = \delta_k d_{k+1}^{\text{LS}} + \gamma_k d_{k+1}^{\text{FR}} + (1-\delta_k-\gamma_k) d_{k+1}^{\text{PRP}}. \tag{14}$$

Multiplying (14) by g_{k+1}^T from the left, we get

$$g_{k+1}^T d_{k+1} = \delta_k g_{k+1}^T d_{k+1}^{\text{LS}} + \gamma_k g_{k+1}^T d_{k+1}^{\text{FR}} + (1-\delta_k-\gamma_k) g_{k+1}^T d_{k+1}^{\text{PRP}}. \tag{15}$$

We have proven seven cases.

Case 1. If $\delta_k = 1$ and $\gamma_k = 0$, then we get

$$g_{k+1}^T d_{k+1} = g_{k+1}^T d_{k+1}^{\text{LS}}.$$

We are going to prove that the sufficient descent condition holds for LS, we have

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

In addition, we have

$$\sigma g_k^T d_k \leq y_k^T d_k \implies -\sigma g_k^T d_k \geq -y_k^T d_k \tag{16}$$

where σ is positive, so

$$\begin{aligned} g_{k+1}^T d_{k+1}^{\text{LS}} &= -\|g_{k+1}\|^2 + \frac{\big(g_{k+1}^T y_k\big)\big(g_{k+1}^T d_k\big)}{-g_k^T d_k}, \\ &\leq -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 y_k^T d_k}{-g_k^T d_k}. \end{aligned}$$

By using (16) in the above inequality, we get

$$g_{k+1}^T d_{k+1}^{\text{LS}} \leq -\|g_{k+1}\|^2 - \frac{1}{\sigma}\|g_{k+1}\|^2.$$

Therefore

$$g_{k+1}^T d_{k+1}^{\text{LS}} \leq -c_1\|g_{k+1}\|^2, \text{ with } c_1 = \left(1 + \frac{1}{\sigma}\right) > 0.$$

Case 2. If $\delta_k = 0$ and $\gamma_k = 1$, the relation (15) becomes

$$g_{k+1}^T d_{k+1} = g_{k+1}^T d_{k+1}^{\text{FR}},$$

under the strong Wolfe line search, the FR method satisfies the sufficient descent condition [16]

$$g_{k+1}^T d_{k+1}^{\text{FR}} \leq -c_2\|g_{k+1}\|^2, \tag{17}$$

where $c_2 > 0$.

Case 3. If $\delta_k = 0$ and $\gamma_k = 0$, the relation (15) becomes

$$g_{k+1}^T d_{k+1} = g_{k+1}^T d_{k+1}^{\mathrm{PRP}}.$$

We are going to prove that the sufficient descent condition holds for PRP when the strong Wolfe line search is used. Thus, we have:

$$g_{k+1}^T d_{k+1}^{\mathrm{PRP}} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{\|g_k\|^2} g_{k+1}^T d_k, \tag{18}$$

we know that

$$\sigma g_k^T d_k \leq y_k^T d_k$$

where σ is positive, multiply both sides by (-1), we have

$$-\sigma g_k^T d_k \geq -y_k^T d_k.$$

Implies that

$$\|g_k\|^2 \geq -\frac{1}{\sigma}\ y_k^T d_k.$$

Then

$$\frac{g_{k+1}^T y_k}{\|g_k\|^2} g_{k+1}^T d_k \leq -\sigma \frac{g_{k+1}^T y_k}{y_k^T d_k} g_{k+1}^T d_k \leq -\sigma\ \|g_{k+1}\|^2,$$

so the equation (18) becomes

$$g_{k+1}^T d_{k+1}^{\mathrm{PRP}} \leq -(1+\sigma)\|g_{k+1}\|^2 \leq -c_3\|g_{k+1}\|^2,$$

where $c_3 > 0$. Then, the proof is completed.

Case 4. If $\delta_k = 0$ and $0 < \gamma_k < 1$, we get

$$g_{k+1}^T d_{k+1} = \gamma_k g_{k+1}^T d_{k+1}^{\mathrm{FR}} + (1-\gamma_k) g_{k+1}^T d_{k+1}^{\mathrm{PRP}} = g_{k+1}^T d_{k+1}^{\mathrm{FRPRP}}.$$

In [9], Djordjević proved that the sufficient descent condition holds

$$g_{k+1}^T d_{k+1}^{\mathrm{FRPRP}} \leq -c_4\|g_{k+1}\|^2,$$

where $c_4 > 0$.

Case 5. If $\gamma_k = 0$ and $0 < \delta_k < 1$. Then

$$g_{k+1}^T d_{k+1} = \delta_k g_{k+1}^T d_{k+1}^{\mathrm{LS}} + (1-\delta_k) g_{k+1}^T d_{k+1}^{\mathrm{PRP}} = g_{k+1}^T d_{k+1}^{\mathrm{LSPRP}}.$$

The sufficient descent condition is fulfilled and mentioned in [1], such that

$$g_{k+1}^T d_{k+1}^{\mathrm{LSPRP}} \leq -c_5\|g_{k+1}\|^2,$$

where $c_5 > 0$.

Case 6. If $1 - \delta_k - \gamma_k = 0$ and $0 < \delta_k, \gamma_k < 1$, then $\gamma_k = 1 - \delta_k$ and the relation (15) becomes

$$g_{k+1}^T d_{k+1} = \delta_k g_{k+1}^T d_{k+1}^{\mathrm{LS}} + (1-\delta_k) g_{k+1}^T d_{k+1}^{\mathrm{FR}} = g_{k+1}^T d_{k+1}^{\mathrm{LSFR}}.$$

Djordjević proved in [11] that d_{k+1}^{LSFR} satisfies the sufficient descent condition for all k, *i.e.* there exists a number $c_6 > 0$, such that

$$g_{k+1}^T d_{k+1}^{\mathrm{LSFR}} \leq -c_6\|g_{k+1}\|^2.$$

Case 7. [9] Now, we are going to prove that the direction satisfies the sufficient descent condition when $\delta_k \in]0,1[$, $\gamma_k \in]0,1[$ and $0 < \delta_k + \gamma_k < 1$, there exist $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ in which $0 < \lambda_1 \leq \delta_k \leq \lambda_3 < 1, 0 < \lambda_2 \leq \gamma_k \leq \lambda_4 < 1$. Now, from (15), we get

$$\begin{aligned} g_{k+1}^T d_{k+1} &\leq \lambda_1 g_{k+1}^T d_{k+1}^{\mathrm{LS}} + \lambda_2 g_{k+1}^T d_{k+1}^{\mathrm{FR}} + (1-\lambda_3-\lambda_4) g_{k+1}^T d_{k+1}^{\mathrm{PRP}} \\ &\leq -\lambda_1 c_1 \|g_{k+1}\|^2 - \lambda_2 c_2 \|g_{k+1}\|^2 - (1-\lambda_3-\lambda_4) c_3 \|g_{k+1}\|^2. \end{aligned}$$

Therefore

$$g_{k+1}^T d_{k+1} \leq -\tau \|g_{k+1}\|^2, \quad \text{with } \tau = [\lambda_1 c_1 + \lambda_2 c_2 + (1-\lambda_3-\lambda_4) c_3] > 0.$$

So, it is proved that d_{k+1} satisfied the sufficient descent condition.

□

2.2. Convergence analysis

The Zoutendijk condition [32] is often utilized to prove the global convergence of the CG method. Moreover, the following lemma shows that the Zoutendijk condition holds for the proposed method under the strong Wolfe conditions of formulas (4) and (5).

Lemma 2. *Suppose that Assumptions 1 and 2 hold. Consider common iterate (2), where d_k is a descent direction and α_k is determined by the strong Wolfe line search (4) and (5). Then, the Zoutendijk condition*

$$\sum_{k\geq 1} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < \infty, \tag{19}$$

holds.

According to the Assumptions 1 and 2, the strong Wolfe conditions and (13), we conclude that α_k, which is obtained in our new method is not equal to zero, *i.e.* there exists a constant $\lambda > 0$ such that

$$\alpha_k \geq \lambda, \text{ for all } k \geq 0. \tag{20}$$

The following theorem gives the global convergence of SCH method.

Theorem 2. *Suppose that Assumptions 1 and 2 hold, let $\{x_k\}$ be generated by SCH algorithm. Then*

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{21}$$

Proof. We prove this theorem by contradiction. Suppose that formula (21) is not true. Then there exists a constant $c > 0$ in which

$$\|g_k\| \geq c, \ \forall k \geq 1. \tag{22}$$

From Theorem 1 it follows that

$$g_k^T d_k \leq -K\|g_k\|^2, \text{ for all } \ K > 0, \tag{23}$$

in the other hand, according to (11), we get

$$\|y_k\| = \|g_{k+1} - g_k\| \leq L\|s_k\| \leq LD, \tag{24}$$

where $D = \max\{\|x-y\|, x, y \in \mathcal{N}\}$ is the diameter of $\mathcal{N}$ and $s_k = x_{k+1} - x_k$.

We have

$$\begin{aligned} d_{k+1} &= -g_{k+1} + \beta_K^{\mathrm{New}} d_k. \\ \|d_{k+1}\| &\leq \|g_{k+1}\| + |\beta_K^{\mathrm{New}}| \|d_k\|. \end{aligned} \tag{25}$$

From (6), we obtain

$$\begin{aligned}
\left|\beta_K^{\mathrm{New}}\right| &\leq \left|\beta_K^{\mathrm{LS}}\right| + \left|\beta_K^{\mathrm{FR}}\right| + \left|\beta_K^{\mathrm{PRP}}\right|, \\
&= \frac{\left|g_{k+1}^T y_k\right|}{\left|d_k^T g_k\right|} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} + \frac{\left|g_{k+1}^T y_k\right|}{\|g_k\|^2}, \\
&\leq \frac{\|g_{k+1}\|\,\|y_k\|}{K\|g_k\|^2} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} + \frac{\|g_{k+1}\|\,\|y_k\|}{\|g_k\|^2}, \\
&\leq \frac{\Gamma L D}{K c^2} + \frac{\Gamma^2}{c^2} + \frac{\Gamma L D}{c^2} = M,
\end{aligned}$$

where the first inequality follows from $0 < \delta_k, \gamma_k < 1$ and $1 - \delta_k - \gamma_k < 1$, the second inequality applies the Cauchy–Schwarz inequality and (23), the last inequality uses (12), (22) and (24).

Thus, it follows from (3) and (20) that

$$\|d_{k+1}\| \leq \|g_{k+1}\| + |\beta_K^{\mathrm{New}}|\ \|d_k\| \leq \|g_{k+1}\| + \frac{|\beta_K^{\mathrm{New}}|\|s_k\|}{\alpha_k} \leq \Gamma + \frac{MD}{\lambda} = W,$$

which implies that

$$\sum_{k\geq 1} \frac{1}{\|d_k\|^2} = +\infty.$$

On the other hand, from (19), (22) and (23), we get

$$K^2 c^4 \sum_{k\geq 1} \frac{1}{\|d_k\|^2} \leq \sum_{k\geq 1} \frac{K^2\|g_k\|^4}{\|d_k\|^2} \leq \sum_{k\geq 1} \frac{(d_k^T g_k)^2}{\|d_k\|^2} < +\infty \tag{26}$$

which contradicts Lemma 2. Hence, equation (22) does not hold, and the claim (21) is proved. □

3. Numerical analysis

This section is devoted to testing the implementation of the new method. Based on this, we compare the computational performance of the proposed method with some known algorithms, such as the LS, FR, PRP conjugate gradient algorithms, and the new hybrid methods: FRPRPCC (Fletcher–Reeves–Polak–Ribiere–Polyak–conjugate-condition) from [9] which we call DJA here, and hFRPRPDY from [17] which we call HYB here. For this comparisons, we consider 400 unconstrained optimization test problems from the CUTE library [7] along with other large-scale optimization problems presented in [3]. Above and beyond, we selected 30 large-scale unconstrained optimization problems in extended or generalized form. More to the point, each problem is tested for several variables: $n = 2, 4, \ldots, 25000$. Nonetheless, the analysis was based on the number of iterations and central processing unit CPU time. For the numerical tests, the iterations are terminated when $\|g_k\|_\infty < 10^{-6}$, at which $\|\cdot\|_\infty$ is the maximum absolute component of a vector, the parameters in the strong Wolfe line searches are chosen to be $\delta = 10^{-3}$ and $\sigma = 10^{-4}$ and the hybridization parameter $\gamma_k = 0.5$. On the other hand, all programs are written in Matlab and compiler settings on the PC machine with Intel(R) Core(TM) i3-4030U CPU @1.90 GHz processor and 4GB RAM and Windows seven professional system.

Comparisons of these methods are given on the following two sides. On the first side, for the ith problem, let f_i^{M1} and f_i^{M2} be the optimal value found by $M1$ method and $M2$ method, respectively. We say that, for the particular problem ith, the performance of the $M1$ method was better than the performance of the $M2$ method if

$$\left|f_i^{M1} - f_i^{M2}\right| < 10^{-3}, \tag{27}$$

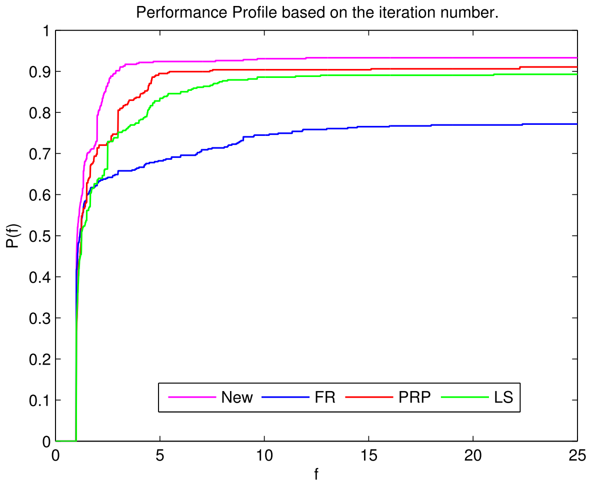

FIGURE 1. Performance Profile based on the iteration number. SCH versus LS, FR and PRP conjugate gradient algorithms.

and number of iterations, or CPU time of $M1$ method is less than those of $M2$ method, respectively. On the other side, to obtain complete comparisons in CPU time, we used the profile of Dolan and Moré [12] to evaluate and compare the performance of the set of methods S on a test set P. Assume that S consists of n_s methods, P consists of n_p problems. For each problem, $p \in P$ and method $s \in S$ denote $t_{p,s}$ be the computing time required to solve problem p by method s. The comparison between different methods is based on the performance ratio defined by $r_{p,s} := t_{p,s} | \min_{s \in S} t_{p,s}$. In consequence, the performance profile is given by

$$\rho_s(\tau) = \frac{1}{n_p} \text{size}\{p \in P : \log_2 r_{p,s} \leq \tau\}, \ \ \forall \tau \in \mathbb{R}^+,$$

where $\rho : \mathbb{R} \to [0,1]$ and $1 \leq s \leq n_s$. The function ρ_s is the distribution function for the performance ratio. Moreover, ρ_s for a method is a non-decreasing, piecewise constant function, continuous from the right at each breakpoint. Note that $\rho_s(\tau)$ is the probability for method $s \in S$ that $\log_2 r_{p,s}$ is within a factor $\tau \in \mathbb{R}^+$ of the best possible ratio. Obviously, when τ takes a certain value, a method with a high value of $\rho_s(\tau)$ is preferable or represents the best method.

In the first set of numerical experiments, we compare the performance of our new algorithm to the LS, FR and PRP conjugate gradient algorithms. Figures 1 and 2 represent the performance profiles of the new method *versus* LS, FR and PRP based on the CPU time and number of iterations, respectively. The two figures show that the new method is superior to the other conjugate gradient methods on the testing problems.

In the second set of numerical experiments, we present a comparison with the new hybrid method FRPRPCC from [9], which we call DJA here, this comparison shall be under the same above conditions. Moreover, Figures 3 and 4 represent the performance profiles of SCH *versus* DJA based on the number of iterations and CPU time.

The two figures show that our method performs better than the DJA method for the number of iterations and CPU time. Simultaneously, according to the performance of DJA method in [9], we can conclude that our method is also better than CCOMB from [4], HYBRID from [2], the algorithm of Touatti–Ahmed and Storey (ToAhS) from [26], the algorithm of Hu and Storey (HuS) from [19] and the algorithm (GN) of Gilbert and Nocedal from [15].

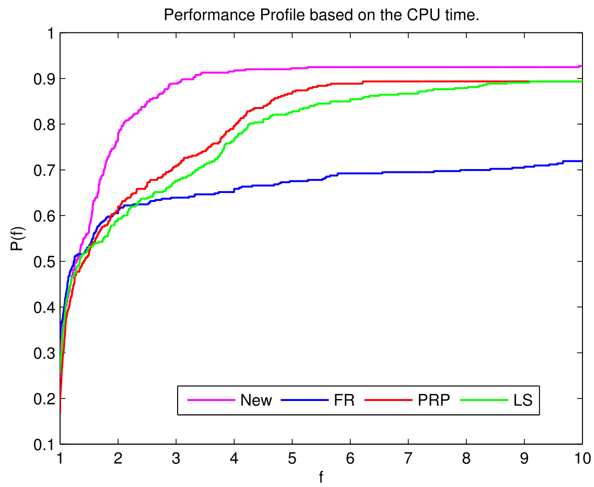

FIGURE 2. Performance Profile based on the CPU time. SCH versus LS, FR and PRP conjugate gradient algorithms.

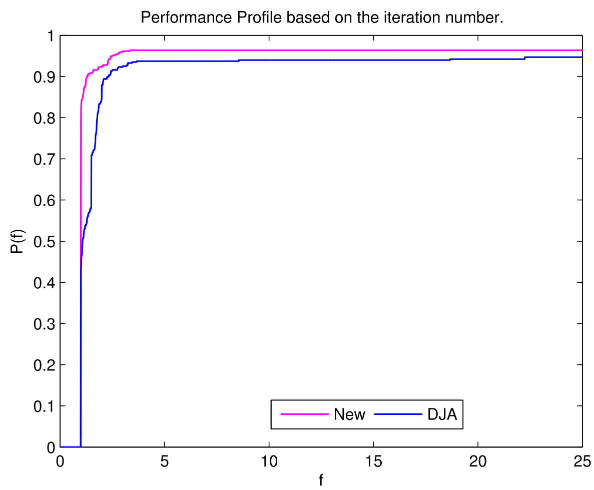

FIGURE 3. Performance Profile based on the iteration number. SCH versus hybrid FRPRPCC (DJA).

The third set of numerical experiments compares our new method to the hybrid conjugate gradient algorithm hFRPRPDY from [17], which we call HYB here. Besides, Figures 5 and 6 represent the performance profiles of the new method *versus* HYB based on the CPU time and number of iterations, respectively.

From these figures illustrated above, we notice that the new algorithm behaves similarly to or better than the HYB method inspired by it.

4. Application of the New CG method to regression analysis

Novel coronavirus-19 (COVID-19) is a new chain of corona group viruses that was not recognized in human history earlier than December 2019. It was first discovered in Wuhan, China [30] and has spread to various urban areas in China as well as approximately 196 different countries of the world. It has since been declared

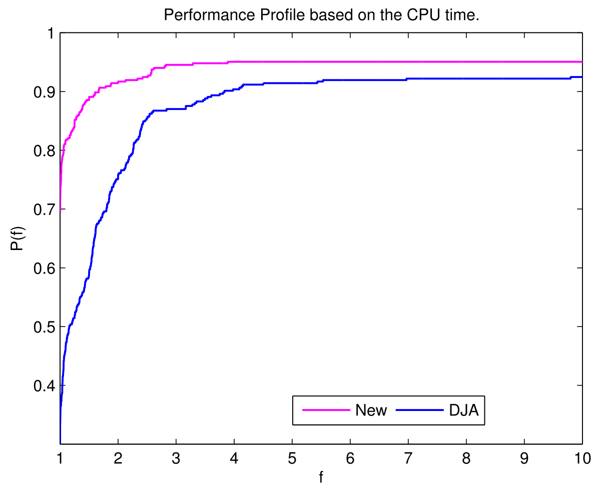

FIGURE 4. Performance Profile based on the CPU time. SCH versus hybrid FRPRPCC (DJA).

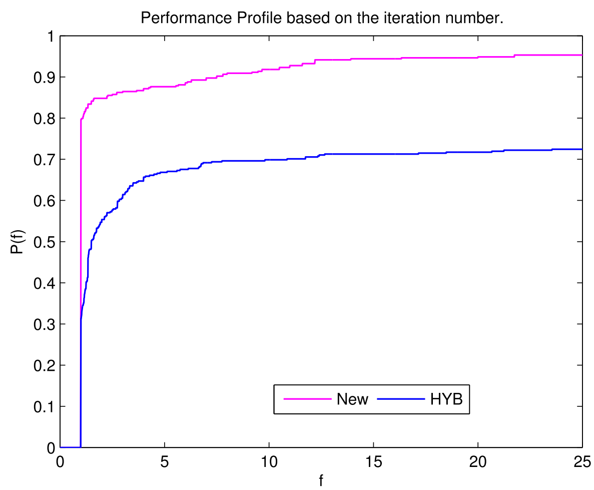

FIGURE 5. Performance Profile based on the iteration number. SCH versus hybrid hFRPRPDY (HYB).

an outbreak by the World Health Organization (WHO). It is difficult to take a single point of view on this virus's origin. It can be due to a seafood market exchange, the people's migration from one location to another, or the transmission from animals to humans. Most people infected by the virus will develop mild to moderate symptoms, such as mild fever, cold, and difficulty breathing, and recover without special treatment. According to data reported by the WHO, on the 20th of October 2020, the laboratory declared that the number of confirmed cases is over 40 million, with more than one million deaths recorded in 215 regions and countries around the world since the disease was first reported in Wuhan.

Mathematical modeling plays a vital role in describing the epidemic of infectious diseases and thus overcoming the same at an early stage. Recently, numerous studies modeled various aspects of the coronavirus outbreak, and the application of numerical methods on some COVID-19 models was also studied [25, 31].

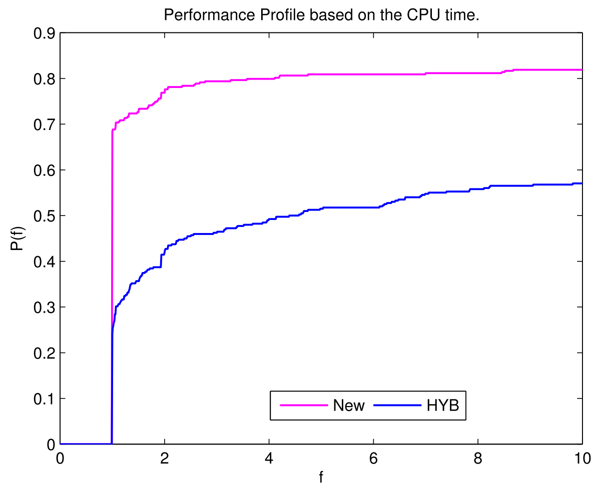

FIGURE 6. Performance Profile based on the CPU time. SCH versus hybrid hFRPRPDY (HYB).

Besides, this paper aims to investigate the performance of the proposed method on a parameterized COVID-19 regression model. For deriving the COVID-19 regression model, the study will consider the total confirmed cases of the infection from January 2020 until September 2020. Subsequently, the obtained data would be transformed into an unconstrained optimization problem, which would later be solved using the proposed method.

Regression analysis is one of the most effective statistical tools for modeling problems in the applied sciences, physical sciences, management and many others. Based on the previous description, we can describe regression analysis as a statistical technique used to estimate the relationship between a dependent variable and one or more independent variables. In virtue of this, the function of regression analysis is defined as follows:

$$y = h(x_1, x_2, \ldots, x_p + \epsilon), \tag{28}$$

where $x_i, i = 1, 2, \ldots, p, p > 0$ is the predictor, y is the response variable, and ϵ is the error. For any problem related to regression analysis, the linear regression function can be derived by computing y such that:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_p x_p + \epsilon, \tag{29}$$

with $a_0, \ldots, a_p$ representing the regression parameters, these parameters are estimated to minimize the error ϵ value. This scheme is often used when the relationship between x and y is approximated by a straight line. However, these cases rarely occur because most problems are often nonlinear. Therefore, the nonlinear regression scheme is frequently used. In this paper, we considered the nonlinear regression one.

To derive the approximate function, we consider the data from the global confirmed cases of COVID-19 from January to September 2020. Table 1 illustrates the process description from the statistics obtained from the World Health Organization [29]. We have data for nine months (Jan–Sept), the months of data collection would be denoted by x-variable, and the y-variable would denote the confirmed cases corresponding to these months. However, the data for eight months (Jan–Aug) would be considered for fitting the data, while the data for September 2020 would be reserved for error analysis.

From the above data, the approximate function for the nonlinear least square method is defined by:

$$f(x) = -25932 + 14512x + 3294.5x^2. \tag{30}$$

The above function (30) will be utilized when approximating the y data values based on x data values from January to August. Let x_j denote the number of months and y_j be the confirmed cases for that month.

TABLE 1. Statistics of confirmed cases of COVID-19, January–Sepember, 2020.

Monthly data (Jan–Sept) (x)	Data of confirmed COVID-19 cases (y)	Statistics of COVID-19 in %
1	2010	0.16
2	1852	0.14
3	58 863	4.7
4	74 019	6.0
5	115 577	9.3
6	172 158	13.9
7	293 238	23.6
8	269 338	21.7
9	254 423	20.5

TABLE 2. Test results for optimization of quadratic model for SCH, LS, FR and PRP.

Initial points	SCH		LS		FR		PRP	
	NOI	CPU	NOI	CPU	NOI	CPU	NOI	CPU
$(2, 2, 2)$	244	2.0670	408	2.3450	F	F	F	F
$(3, 3, 3)$	471	2.9430	519	3.1970	F	F	252	1.4530
$(10, 10, 10)$	410	2.7200	429	2.5140	F	F	F	F
$(13, 13, 13)$	632	4.4500	691	4.2880	F	F	F	F
$(30, 30, 30)$	344	2.1830	457	2.7360	F	F	F	F

Based on this information, the above least squares method (30) is transformed into the following unconstrained minimization problems.

$$\min_{x\in\mathbb{R}^n} f(x) = \sum_{j=1}^{n}\left(\left(u_0 + u_1 x_j + u_2 x_j^2\right) - y_j\right)^2. \tag{31}$$

The data of the first eight months from Table 1 are utilized to formulate the nonlinear quadratic model for the least square method, which is further used to derive the unconstrained optimization model. Based on the above discussion, it is evident that there exist some parabolic relations between the data x_j and the value of y_j with the regression function defined by (30) and the regression parameters u_0, u_1 and u_2

$$\min_{x\in\mathbb{R}^2} \sum_{j=1}^{n} E_j^2 = \sum_{j=1}^{n}\left(\left(u_0 + u_1 x_j + u_2 x_j^2\right) - y_j\right)^2. \tag{32}$$

Subsequently, using the data of Table 1, we transform (32) to obtain our nonlinear quadratic unconstrained minimization model as follows:

$$8u_1^2+72u_1u_2+408u_1u_3-1974110u_1+204u_2^2+2592u_2u_3-12593164u_2+8772u_3^2-84833792u_3+210479037915. \tag{33}$$

The above nonlinear quadratic model was constructed using data from January to August. Meanwhile, the data for September is reserved for relative error analysis of the predicted data. At present, we can apply the SCH, LS, FR and PRP methods for solving the model (33) under the strong Wolfe line search conditions (4) and (5), we obtain the performance results based on iteration numbers and CPU time illustrated in Table 2.

To overcome the difficulty of computing the values of u_0, u_1, u_2 using matrix inverse, we implement the previously mentioned methods using different initial points. We terminate the computation i:

- The defined stopping criteria are satisfied based on the value defined for each function.
- The method is unable to solve the model.

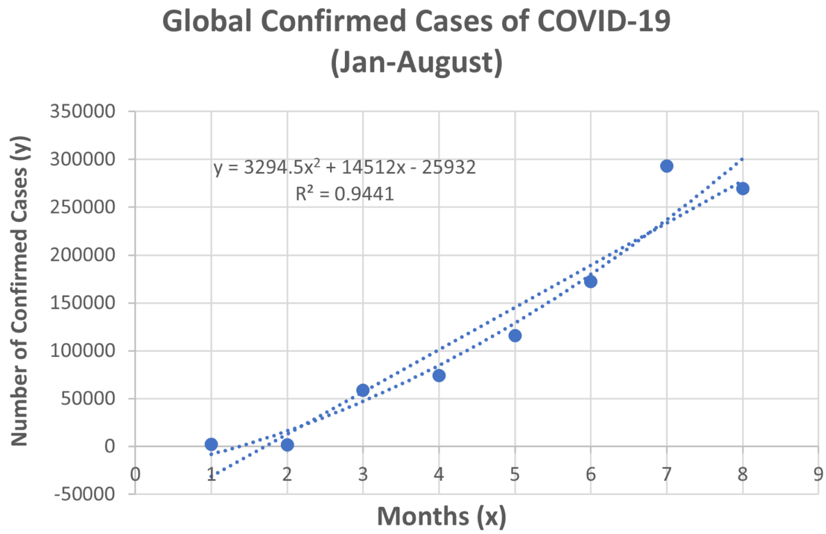

FIGURE 7. Nonlinear quadratic trend line for confirmed cases of COVID-19.

4.1. Trend line method

In this subsection, we aim to estimate the confirmed cases of COVID-19 for eight (8) months using the proposed SCH, some known CG, and least square methods. From the actual data obtained from Table 2, we use Microsoft Excel software to plot the trend line as demonstrated in Figure 7. Furthermore, to show the efficiency of the proposed method, we compare the approximation functions of the SCH method with the functions of the LS, FR, PRP and trend line methods. Based on the results illustrated in Table 2, it is evident that the suggested SCH method is faster and more efficient compared to the used methods. On the other hand, from the plot, it is clear that the trend line equation obtained is a nonlinear quadratic equation. In light of this, the ideal purpose of the regression analysis is to estimate $a_0, a_1, \dots, a_p$ where the error ϵ is minimized. From the above discussion, we can conclude that the SCH method can be used as an alternative to the trend line method and the least squares method, which implies that the method is applicable to real-world situations.

5. CONCLUSION

In light of the facts above, the hybrid CG methods are usually obtained based on the classical CG methods by integrating their advantages. In this paper, we proposed a new hybrid conjugate gradient algorithm in which the famous parameter β_k is computed as a convex combination of $\beta_k^{\mathrm{LS}}, \beta_k^{\mathrm{FR}}$ and β_k^{PRP} algorithms. Based on some conditions, we show that the proposed algorithm enjoys sufficient descent condition and converges globally under strong Wolfe line search. Further, numerical experiments are considered to illustrate the performance of the proposed method. The results show that the new method is more effective and has a better convergence rate than LS, FR, PRP, FRPRPCC and hFRPRPDY methods. More to the point, our proposed method can solve the COVID-19 case model.

Acknowledgements. We want to thank the professors of Laboratory Informatics and Mathematics (LIM) (University of Souk Ahras, Algeria) who have always been helping and giving us advice.

REFERENCES

[1] A.B Abubakar, P. Kumam, M. Malik and A.I Hassan, A hybrid conjugate gradient based approach for solving unconstrained optimization and motion control problems. *Math. Comput. Simul.* **201** (2022) 640–657.

[2] N. Andrei, A hybrid conjugate gradient algorithm for unconstrained optimization as a convex combination of Hestenes–Stiefel and Dai–Yuan. *Stud. Inf. Control* **17** (2008) 55–70.

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

[4] N. Andrei, New hybrid conjugate gradient algorithms for unconstrained optimization. *Encycl. Optim.* (2009) 2560–2571.

[5] N. Andrei, Another nonlinear conjugate gradient algorithm for unconstrained optimization. *Optim. Methods Softw.* **24** (2009) 89–104.

[6] N. Andrei, Nonlinear Conjugate Gradient Methods for Unconstrained Optimization. Springer (2020).

[7] I. Bongartz, A.R. Conn, N. Gould and P.L. Toint, CUTE: constrained and unconstrained testing environments. *ACM Trans. Math. Softw. (TOMS)* **21** (1995) 123–160.

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

[9] S.S. Djordjević, New hybrid conjugate gradient method as a convex combination of FR and PRP methods. *Filomat* **30** (2016) 3083–3100.

[10] S.S. Djordjević, New hybrid conjugate gradient method as a convex combination of HS and FR conjugate gradient methods. *J. Appl. Math. Comput.* **2** (2018) 366–378.

[11] S.S. Djordjević, New hybrid conjugate gradient method as a convex combination of LS and FR conjugate gradient methods. *Acta Math. Sci.* **39** (2019) 214–228.

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

[13] R. Fletcher, Practical Methods of Optimization, Unconstrained Optimization. Vol. 1. John Wiley and Son, New York (1980).

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

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

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

[17] S.B. Hanachi, S. Badreddine and M. Belloufi, New iterative conjugate gradient method for nonlinear unconstrained optimization. *RAIRO: Oper. Res.* **56** (2022) 2315–2327.

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

[19] Y.F. Hu and C. Storey, Global convergence result for conjugate gradient method. *J. Optim. Theory Appl.* **71** (1991) 399–405.

[20] S. Li and Z.B. Sun, A new hybrid conjugate gradient method and its global convergence for unconstrained optimization. *Int. J. Pure Appl. Math.* **63** (2010) 84–93.

[21] A.J. Liu, and S. Li, New hybrid conjugate gradient method for unconstrained optimization. *Appl. Math. Comput.* **245** (2014) 36–43.

[22] D. Liu and C. Story, Efficient generalized conjugate gradient algorithms, part 1: theory. *J. Optim. Theory App.* **69** (1991) 129–137.

[23] E. Polak and G. Ribiere, Note sur la convergence de méthodes de directions conjuguées. *Revue Française D'informatique et De Recherche Opérationnelle, Série Rouge.* **3** (1969) 35–43.

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

[25] I.M. Sulaiman, N.A. Bakar, M. Mamat, B.A. Hassan, M. Malik and A.M. Ahmed, A new hybrid conjugate gradient algorithm for optimization models and its application to regression analysis. *Indones. J. Electr. Eng. Comput. Sci.* **23** (2021) 1100–1109.

[26] D. Touati-Ahmed and C. Storey, Efficient hybrid conjugate gradient techniques. *J. Optim. Theory Appl.* **64** (1990) 379–397.

[27] P. Wolfe, Convergence conditions for ascent methods. *SIAM Rev.* **11** (1969) 226–235.

[28] P. Wolfe, Convergence conditions for ascent methods. II: some corrections. *SIAM Rev.* **13** (1971) 185–188.

[29] World heath organization, Report on coronavirus (COVID-19) (2020).

[30] World Health Organization (WHO), Naming the coronavirus disease (COVID-19) and the virus that causes it. *Braz. J. Implantol. Health Sci.* **2** (2020).

[31] T.J. Wu, K. Leung and M.G. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* **395** (2020) 689–697.

[32] G. Zoutendijk, Nonlinear programming, computational methods. *Integer Nonlinear Program.* (1970) 37–86.