

Global convergence of new conjugate gradient method with inexact line search

Chergui Ahmed¹, Bouali Taher²

¹Mathematics and Computer Science Department, Laboratory of Mathematics, Informatics and Systems (LAMIS), Tebessa University, Algeria

¹Department of Mathematics, Khenchela University, Khenchela, Algeria

²Department of Mathematics, Faculty of Science, Jazen University, Jazen, Saudi Arabi

Article Info

Article history:

Received Feb 13, 2020

Revised Aug 13, 2020

Accepted Oct 1, 2020

Keywords:

Conjugate gradient

Global convergence

Strong wolf line search

Unconstrained optimization

ABSTRACT

In this paper, We propose a new nonlinear conjugate gradient method (FRA) that satisfies a sufficient descent condition and global convergence under the inexact line search of strong wolf powell. Our numerical experiment shaw the efficiency of the new method in solving a set of problems from the CUTEst package, the proposed new formula gives excellent numerical results at CPU time, number of iterations, number of gradient ratings when compared to WYL, DY, PRP, and FR methods.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Chergui Ahmed

Mathematics and Computer Science Department

Laboratory of Mathematics, Informatics and Systems (LAMIS)

Larbi Tebessi University-Tebessa, Algeria

Department of Mathematics, Khenchela University, Khenchela 40004, Algeria

Email: ahmed.chergui@univ-tebessa.dz

1. INTRODUCTION

The optimization problem finds application in several fields, such as pure mathematics, mathematical and computational physics, mathematical physics, fluid dynamics, an traffic routing in telecommunication systems [1], cyber-physical security [2], intelligent transportation systems [3], and smart grids [4]. The conjugate gradient method is an effective one for solving large-scale unconstrained optimization problems because it need not the storage of any matrices. Well-known conjugate gradient methods are [5-9]. Global convergence properties of these methods have been studied [9-12].

In this paper, we consider the following unconstrained optimization problem:

$$(p): \min\{f(x)\}: x \in \mathbb{R}^n \quad (1)$$

where f smooth and its gradient $\nabla f(x_k)$ is available

$$x_{k+1} = x_k + t_k d_k, t_k > 0 \quad k = 0; 1; 2; 3 \quad (2)$$

where t_k a positive step size along the search direction obtained by line search. x_k is the current iterative point and d_k is search direction has the form

$$\text{By } d_k = \begin{cases} -g_k & \text{if } k = 1 \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 2 \end{cases} \quad (3)$$

where β_k a parameter characterizes the CG method and g_k denotes $\nabla f(x_k)$;

The main difference among CG methods is in the formulas of computing their parameters. Some of the well known CG methods are reviewed in [13]. A very famous formula for computing g_k is proposed by Fletcher and Reeves (FR) [5] as following

$$B_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \text{ Fletcher Reeves [5]} \quad (4)$$

$$B_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} \text{ Polak - Ribière - Polyak [6]} \quad (5)$$

$$B_k^{DY} = \frac{g_k^T g_k}{((g_k - g_{k-1})^T d_{k-1})} \text{ Dai - Yuan [9]}, \quad (6)$$

$$B_k^{WYL} = \frac{g_k^T g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}}{g_{k-1}^T g_{k-1}} \text{ Wei et al. [10]}, \quad (7)$$

where $\|\cdot\|$ Denotes the Euclidean norm. This formula is usually considered the .rst nonlinear CG parameter [14]. Having the direction d_k , the ideal choice for the steplength t_k would be the global minimizer of, conditions that require t_k satisfying. In order to find the step length (α_k), we use strong wolf powell (SWP) line search,

$$f(x_k + t_k d_k) \leq f(x_k) + \delta t_k g_k^T d_k \quad (8)$$

$$|\nabla f(x_k + t_k d_k)^T \cdot d_k| \leq -\sigma g_k^T d_k \quad (9)$$

where $(0 < \delta < \frac{1}{2})$ and $(0 < \sigma < 1)$

Strong Wolfe conditions used for establishing the global convergence in [9, 12], and [14-17]. The pioneer works about the global convergence of FR method with inexact line search was proposed by Al-Baali [18]. He proved that the FR method satisfied the sufficient descent directions and globally convergent under the (SWP) conditions with $0 < \delta < \sigma < \frac{1}{2}$, in [9, 19]. This result was extended to $\sigma = \frac{1}{2}$. It is shown that FR method with the (SWP) line search may not be a descent direction for the case that $\sigma > \frac{1}{2}$. For the B_k^{FR} , neither Armijo nor Wolfe line search, guarantee that the condition sufficient descent. In 2006 [20], Nocedal, J. and Wright, S. (2006) [21-23]. The paper is organized as follows, in section 2. We introduce the new algorithm for B_k in section3, we analyze the global convergence property of the new method. Finally, numerical results and conclusion in sections 4 and 5.

2. NEW ALGORITHM OF FRA

We propose a new B_k for the CG method. The sequence of iteration x_k in the new method is obtained from (2) for which the direction d_k is computed by (3). While the parameter B_k parameter B_k in the new method is;

$$B_k^{FRA} = \lambda \frac{\|g_k^T\|^2}{\|g_{k-1}^T\|^2} \lambda \in (0; 1) \quad (10)$$

where FRA designed the new modified method by Ahmed Chergui.

Note that, for the direction d_k defined by (3), with the CG parameter computed by (10), we have,

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

By the Cauchy-Schwarz inequality, it can be concluded that,

$$g_k^T d_k \leq -\|g_k\|^2 + \lambda \|g_k\|^2 = (-1 + \lambda) \|g_k\|^2 < 0 \tag{12}$$

So, the new direction d_k is satisfied.

In the new CG method, the step t_k is determined by the (SWP). To this aim, we use a backtracking approach to compute the steplength. Now we are ready to propose the algorithm of the new CG method (10)

Algorithm 1

- Step 1: Given $x_0 \in R^n$ set $k = 1$. $\epsilon \in (0, 1)$ set $d_0 = -g_0 = -\nabla f(x_0)$
- Step 2: Compute B_k by (10), (4), (5), (6); (7)
- Step 3: Compute d_k by (3); if $\|g_k\| = 0$, then stop.
- Step 4: Calculate step length t_k by (8) and (9) line search, $\sigma = 0.1$, $\delta = 0.01$
- Step 5: Let $x_{k+1} = x_k + t_k d_k$.
- Step 6: if $f(x_k) < f(x_{k-1})$ and $\|\nabla f(x_k)\| < \epsilon$, then stop,
Otherwise Set $k = k + 1$ go to step 2

3. THE GLOBAL CONVERGENCE PROPRIETES

In this section, we analyze the convergence of FRA method. To this aim, we made the following assumption:

Assumption 1

- (H1) The objective function f is bounded below on the level set R^n and is continuous and differentiable in neighborhood V of the level set $\Omega = \{x \in R^n; f(x) < f(x_0)\}$
- (H2) The gradient g_k is Lipschitz continuous in V , so a constant $M \geq 0$ exists, such that

$$\|g(x) - g(y)\| \leq M \|x - y\| \text{ For all } x, y \in V \tag{13}$$

The following lemma provides a lower bound for the steplength t_k (generated by Algorithm 1). The result of this lemma will be needed in the rest of this section.

3.1. Sufficient descent condition

Theorem 1: suppose that the sequence $\{g_k\}$ and $\{d_k\}$ are generated by (2) (3) and FRA .the step length t_k , is determined by inexact line search (9) and (10) if $g_k \neq 0$, then d_k possesses the sufficient descent condition: $g_k^T d_k \leq -C \|g_k\|^2$

Proof: By the formula (10), we have the following: $B_k^{FRA} = \lambda \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \geq 0$

$$\text{Hence we obtain } 0 \leq B_k^{FRA} \leq \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \tag{14}$$

Using (9) and (14), we get,

$$|B_k^{FRA} g_k^T \cdot d_{k-1}| \leq \sigma \frac{\|g_k\|^2}{\|g_{k-1}\|^2} |g_k^T \cdot d_{k-1}| \tag{15}$$

By (3), we have $d_k = -g_k + \beta_k d_{k-1}$

$$\frac{g_k^T \cdot d_k}{\|g_k\|^2} = -1 + B_k^{FRA} \frac{g_k^T d_{k-1}}{\|g_k\|^2} \tag{16}$$

we have $g_0^T d_0 \leq -\|g_0\|^2 < 0$

If $g_0 \neq 0$; suppose that $d_i; i = 1, 2, \dots, k;$ are all descente directions, that is $g_k^T d_k < 0$

By (16); we get;

$$|B_k^{FRA} g_k^T \cdot d_{k-1}| \leq -\sigma \frac{\|g_k\|^2}{\|g_{k-1}\|^2} g_k^T \cdot d_{k-1} \tag{17}$$

That is;

$$\sigma \frac{\|g_k\|^2}{\|g_{k-1}\|^2} g_k^T \cdot d_{k-1} \leq B_k^{FRA} g_k^T \cdot d_{k-1} \leq -\sigma \frac{\|g_k\|^2}{\|g_{k-1}\|^2} g_k^T \cdot d_{k-1} \tag{18}$$

As shown in (17) and (18) deduce $-1 + \sigma \frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -1 - \sigma \frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}$
 By repeating this process and the fact $g_0^T d_0 = -\|g_0\|^2$, we have

$$-\sum_{i=0}^{k-1} (\sigma)^i \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -2 + \sum_{i=0}^{k-1} (\sigma)^i \quad (19)$$

As shown in (19). Can be written as;

$$-\frac{1}{1-\sigma} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -2 + \frac{1}{1-\sigma} \quad (20)$$

By making the restriction $\sigma \in (0, 0.1)$ we have $g_k^T d_k < 0$.
 Now, we prove the sufficient descent condition of d_k if $\sigma \in (0, 1)$
 Set $c = -2 + \frac{1}{1-\sigma}$ then $0 < c < 1$, and (17) turns out to be;

$$c - 2 \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -c \quad (21)$$

Thus we obtain $g_k^T d_k \leq -C \|g_k\|^2$ Or $C = -2 + \frac{1}{1-\sigma}$.

3.2. Convergent analysis

Lemma 1 Let the step length t_k is generated by Algorithm 1. Then, under the assumptions H1 and H2, there is a positive constant C such that,

$$t_k \geq C \frac{\|g_k\|^2}{\|d_k\|^2} \quad (22)$$

Proof: Subtracting $g_k^T d_k$ from both sides of (10) and using (19) we have

$$-(1-\sigma)g_k^T d_k \leq (g_{k+1} - g_k)^T \leq M t_k \|d_k\|^2 \quad (23)$$

therefore;

$$t_k \geq -\frac{(1-\sigma) g_k^T d_k}{M \|d_k\|^2} \quad (24)$$

with (10) we obtain:

$$t_k \geq -\frac{(1-\sigma) \|g_k\|^2}{M \|d_k\|^2} \quad (25)$$

This inequality means that (25) satisfies with $C = -\frac{(1-\sigma)}{M}$, the proof is completed. The next lemma is known as Zoutendijk condition [24].

Lemma 2: Suppose assumption 1 hold and d_k is generated by Algorithm 1, then;

$$\sum_{n=0}^{\infty} \frac{\|g_n\|^4}{\|d_n\|^2} < \infty \quad (26)$$

Proof: From (10) for any k we have;

$$f(x_k) - f(x_k + t_k d_k) \geq -\delta t_k g_k^T d_k \geq -\delta \frac{(1-\sigma) (g_k^T d_k)^2}{M \|d_k\|^2} \quad (27)$$

Moreover, from the hypothesis (1), we have that $\{f(x_k)\}$ is a decreasing sequence and has a limit in, which shows that $\lim_{k \rightarrow \infty} f(x_{k+1}) < +\infty$ and after (28) we have;

$$+\infty > f(x_1) - \lim_{k \rightarrow \infty} f(x_{k+1}) \geq \delta \frac{(1-\sigma)}{M} \sum \frac{(g_k^T d_k)^2}{\|d_k\|^2} \tag{28}$$

Then $\sum \frac{(g_k^T d_k)^2}{\|d_k\|^2} \leq +\infty$, so, the proof is completed.

Theorem 2: we assume that H1, H2 hold, and the sequence $\{x_k\}$ is generated by the Algorithm 1, then, $\lim_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0$

4. NUMERICAL EXPERIMENT

In this part, we report numerical experiments that indicate the efficiency of the new algorithm. To this aim, we implement the new algorithm (Algorithm 1), Fletcher and Reeves (FR) algorithm and the modified Fletcher and Reeves (FR), WYL [10], DY [9], PRP [6]. The numerical results are given in the different initial points. We considered $\varepsilon = 10^{-6}$, $\sigma = 0.1$, $\delta = 0.01$, under inexact line search of (SWP). We used MATLAB R2010 the performance results are shown in Figures 1-5 and compare their results obtained from solving of 17 test problems from [25].

In our experiments the stopping tolerance for the algorithms is Also, a failure is reported when $NI > 20000$ or when the step length t_k become less than $\text{eps}=10^{-6}$. We use we use the performance profiles in [26, 27]. The total number of iterations, the total number of function evaluations, and the running time of each algorithm number of function evaluations. It can be seen that the FRA is the best solver with probability around 80%, while the probability of solving a problem as the best solver is around 60%, 26%, 18% and 7% for the FR, PRP, WYL and the DAY respectively. The performance index in. Figure 2 is the total number of iterations. From this figure, we observe that the NEW method (FRA) obtains the most wins on approximately 70% of all test problems an the probability of being best solver is 55%, 29%, 26% and 8% for the FR, PRP, WYL and the DAY respectively.

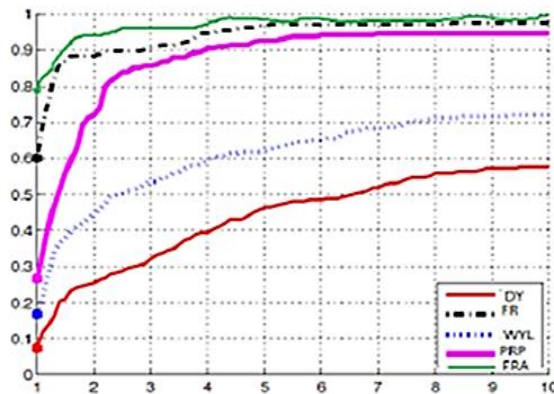


Figure 1. Performance profiles based on the number of function evaluations

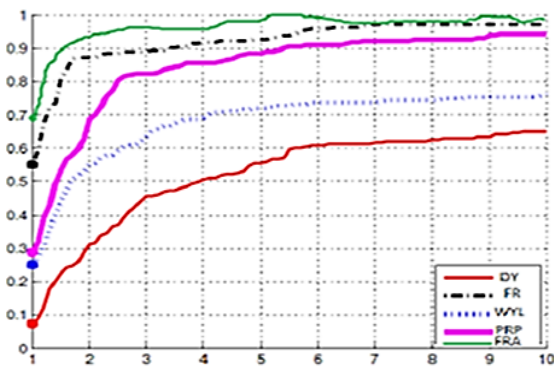


Figure 2. Performance of the number of iterations

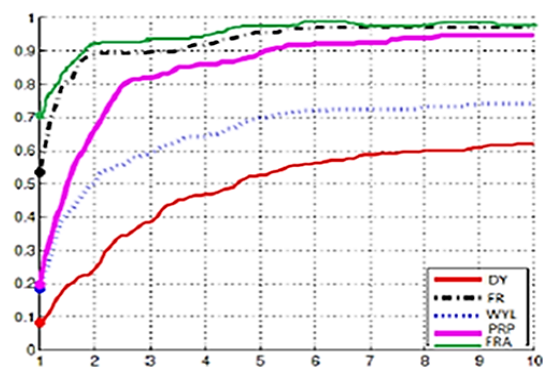


Figure 3. Performance profiles for running times

The CPU time is illustrated in Figure 3. From this figure, it can be observed that the NEW is the best algorithm. Another important factor of these three figures is that the graph of the NEW algorithm grows up faster than the other algorithms. From the presented results, we can observe that the FRA method is best than the FR, PRP, WYL and the DAY methods. In solving unconstrained optimization problems.

Example 1: Extended Rosenbrock function $(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n-1} [100(x_{i-1} - x_i^2)^2 + (x_i - 1)^2]$, $n = 2$, $x_{optimal} = (1, 1)$

In Table 1, The FRA method was successful in all attempts to achieve the optimal solution, while the other methods failed.

- Remark 1: Table 2, shows that “FRA” has the best results since it solves about 100% from the test problems. Figures 4 and 5 list the comparison of FR method and DY, WYL, PRP, FR methods $x_0 = [1, 7]$

Table 1. Numerical results for FRA, FR, PRP, WYL and DY in terms of number iterations (NI) and CPU time with the strong wolf condition $\epsilon = 10^{-6}$; $\sigma = 0.1$; $\delta = 0.01$; $\alpha = 0.9$

Initial point	FRA	FR	PRP	WYL	DY
	NI/CPU	NI/CPU	NI/CPU	NI/CPU	NI/CPU
(10000, 10000)	637/6.41	Failed	Failed	Failed	Failed
(100000, 100000)	934/5.74	Failed	Failed	Failed	Failed
(1000, 1000)	299/0.887	Failed	Failed	Failed	Failed
(-1, 3)	196/0.600	313/2.528	4.66/2.80	14532/78.11	170/2.266
(100, 100)	161/1.68	4693/20.043	Failed	Failed	Failed
(1, 3)	122/0.414	104/4.019	243/1.43	Failed	112/0.23
(0, -9)	163/0.533	355/0.89	340/1.639	Failed	Failed
(1, 7)	67/0.533	230/2.541	737/5.739	4470/15.315	149/1.195

Table 2. Comparing the results obtained in Table 1

Méthod	Ranking	The success rate
FRA	1	100%
FR	2	55%
PRP	3	44%
DY	4	33%
WYL	5	22%

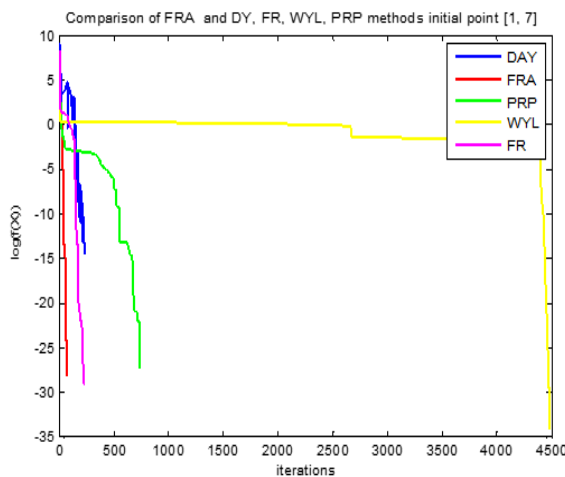


Figure 4. Performance of the number of function evaluations

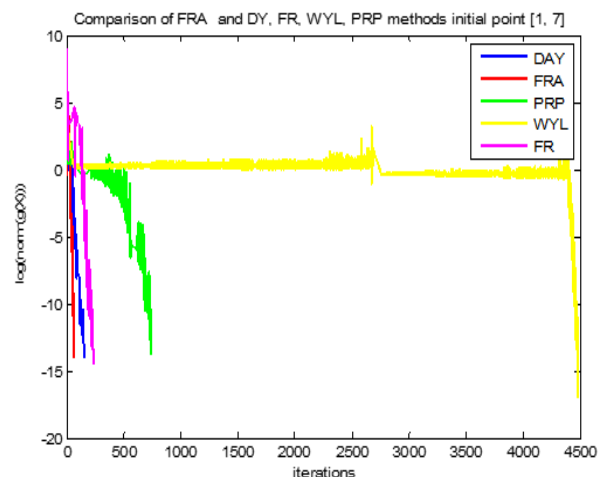


Figure 5. Performance of the number of gradient evaluations

- Remark 2: From the Figures 4 and 5, The FRA method performs better than other methods by selecting a starting point with the Resenbrock function $f(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n-1} [100(x_{i-1} - x_i^2)^2 + (x_i - 1)^2]$, $n = 2$. And she is best performance in terms of values gradients and functions and the number of iterations

5. CONCLUSION

in this paper, we have proposed a new CGmethod named FRA for solving a large-scale unconstrained optimization problem. We proved the global convergence of this method and sufficient descent condition under the inexact line search of (SWP) numerical experiment show that the new method FRA is more efficient than the others methods DAY, WYL, FR, and PRP.

ACKNOWLEDGEMENTS

I would like to thank the anonymous rulers and the editor for their accurate and valuable reading. Suggestions that led to a correction of errors and a significant improvement in the paper.

REFERENCES

- [1] A. Pietrabissa and L. Ricciardi Celsi, "Discrete-Time Selfish Routing Converging to the Wardrop Equilibrium," in *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1288-1294, 2019.
- [2] A. Di Giorgio, et al., "On the optimization of energy storage system placement for protecting power transmission grids against dynamic load altering attacks," *2017 25th Mediterranean Conference on Control and Automation (MED)*, Valletta, 2017, pp. 986-992.
- [3] L. R. Celsi, et al., "On the many-to-many carpooling problem in the context of multi-modal trip planning," *2017 25th Mediterranean Conference on Control and Automation (MED)*, Valletta, 2017, pp. 303-309.
- [4] V. Suraci, L. R. Celsi, A. Giuseppi and A. Di Giorgio, "A distributed wardrop control algorithm for load balancing in smart grids," *2017 25th Mediterranean Conference on Control and Automation (MED)*, Valletta, 2017, pp. 761-767.
- [5] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *The Computer Journal*, vol. 7, pp. 149-154, 1964.
- [6] B. T. Polyak, "The conjugate gradient method in extremum problems," *USSR Computational Mathematics and Mathematical Physics*, vol. 9, no. 4, pp. 94-112, 1969.
- [7] Dai, Y., "Convergence of nonlinear methods," *Journal of Computational Mathematics*, vol. 19, pp. 539-549, 2001.
- [8] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards*, vol. 49, no. 6, pp. 409-436, 1952.
- [9] Y. H. Dai and Y. Yuan, "A nonlinear conjugate gradient method with a strong global convergence property," *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 177-182, 1999.
- [10] Weiet, et al., "The convergence properties of some conjugate gradient methods," *Applied Mathematics and Computation*, vol. 183, pp. 1341-1350, 2006.
- [11] Dai, Y. and Yuan, Y., "Nonlinear conjugate gradient methods," *shanghai scientific*, 2000.
- [12] Gilbert, J. C. and Nocedal, J., "Global convergence properties of conjugate gradient methods for optimization," *SIAM Journal on optimization*, vol. 2, no. 1, pp. 21-42, 1992.
- [13] Guanghui, L., Jiye, H., and Hongxia, Y., "Global convergence of the fletcher-reeves algorithm with inexact linesearch," *Applied Mathematics- A Journal of Chinese Universities*, vol. 10, no. 1, pp. 75-82, 1995.
- [14] Hager, W. W. and Zhang, H., "A survey of nonlinear conjugate gradient methods," *Pacific journal of Optimization*, vol. 2, no. 1, pp. 35-58, 2006.
- [15] Liu, H., Wang, H., Qian, X., and Rao, F., "A conjugate gradient method with sufficient descent property," *Numerical Algorithms*, vol. 70, no. 2, pp. 269-286, 2015.
- [16] Livieris, I. E., Tampakas, V., and Pintelas, P., "A descent hybrid conjugates gradient method based on the memorylessbfgs update," *Numerical Algorithms*, vol. 79, no. 4, pp. 1169-1185, 2018.
- [17] Yao, S., He, D., and Shi, L., "An improved perry conjugate gradient method with adaptive parameter choice," *Numerical Algorithms*, vol. 78, no. 3, pp. 1-15, 2018.
- [18] Al-Baali, M., "Descent property and global convergence of the Fletcher-reeves method with inexact line search," *IMA Journal of Numerical Analysis*, vol. 5, no. 1, pp. 121-124, 1985.
- [19] Hager, W. W. and Zhang, H., "Algorithm 851: Cg descent, a conjugate gradient method with guaranteed descent," *ACM Transactions on Mathematical Software (TOMS)*, vol. 32, no. 1, pp. 113-137, 2006.
- [20] Zhang, L., Zhou, W., and Li, D., "Global convergence of a modified Fletcher-reeves conjugate gradient method with Armijo-type line search," *Numerische Mathematik*, vol. 104, no. 4, pp. 561-572, 2006.
- [21] Nocedal, J. and Wright, S., "Numerical optimization," *Springer Science & Business Media*, 2006.
- [22] S. Sanmtias and E. Vercher E., "A generalized conjugate gradient algorithm," *Journal of Optimization Theory and Applications*, vol. 98, pp. 489-502, 1998.
- [23] Z. J. Shi, "Nonlinear conjugate gradient method with exact line search (in Chinese)," *Acta Mathematica Sinica, English Series*, vol. 24, no. 6, pp. 675-682, 2004.
- [24] G. Zoutendijk, "Nonlinear programming computational methods," in *J. Abadie (Ed.), Integer and Nonlinear Programming, North-Holland*, Amsterdam, pp. 37.86, 1970.
- [25] Gould, N. I., Orban, D., and Toint, P. L., "Cutest: a constrained and unconstrained testing environment with safe threads for mathematical optimization," *Computational Optimization and Applications*, vol. 60, no. 3, pp. 545-557, 2015.
- [26] Dolan, E. D. and Moré, J. J., "Benchmarking optimization software with performance profiles," *Mathematical Programming*, vol. 91, no. 2, pp. 201-213, 2002.
- [27] N. Andrei, "An unconstrained optimization test functions ollection," *Advanced Modeling and Optimization*, vol. 10, pp. 147-161, 2008.