

New Parameter of Conjugate Gradient Method for Unconstrained Nonlinear Optimization

Mohamed Lamine Ouaoa*, Samia Khelladi, Djamel Benterki

*Laboratory of Fundamental and Numerical Mathematics, Department of Mathematics, Faculty of Sciences,
Setif 1 University-Ferhat Abbas, 19000, Algeria*

Abstract We are interested in the performance of nonlinear conjugate gradient methods for unconstrained optimization. In particular, we address the conjugate gradient algorithm with strong Wolfe inexact line search. Firstly, we study the descent property of the search direction of the considered conjugate gradient algorithm based on a new direction obtained from a new parameter. The main objective of this parameter is to improve the speed of convergence of the obtained algorithm. Then, we present a complete study that shows the global convergence of this algorithm. Finally, we establish comparative numerical experiments on well-known test examples to show the efficiency and robustness of our algorithm compared to the algorithm of Hager and Zhang.

Keywords Unconstrained optimization, Conjugate gradient method, Descent direction, Inexact line search, Global convergence

AMS 2010 subject classifications 90C26, 90C30, 65K05

DOI: 10.19139/soic-2310-5070-2069

1. Introduction

Consider the following unconstrained nonlinear optimization problem

$$\begin{cases} \min f(x) \\ x \in \mathbb{R}^n \end{cases}, \quad (1)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable function.

The nonlinear conjugate gradient methods are efficient to solve problem (1), which reflects several concrete problems arising from industry, medicine and engineering, such as image restoration, robotics, sparse signal recovery problems, especially for large dimensions. The iterative scheme of the conjugate gradient method is given as follows:

$$x_1 \in \mathbb{R}^n, \quad x_{k+1} = x_k + \alpha_k d_k, \quad (2)$$

where x_k is the current iterate point, $\alpha_k > 0$ is the step size which can be found by one of the line search methods and d_k is the search direction defined by:

$$d_k = \begin{cases} -g_1 & \text{for } k = 1 \\ -g_k + \beta_k d_{k-1} & \text{for } k \geq 2, \end{cases} \quad (3)$$

where $g_k = \nabla f(x_k)$ is the gradient of f at x_k and β_k is a scalar conjugacy coefficient.

*Correspondence to: Mohamed Lamine Ouaoa (Email: mohamedlamine.ouaoa@univ-setif.dz). Laboratory of Fundamental and Numerical Mathematics, Department of Mathematics, Faculty of Sciences, Setif 1 University-Ferhat Abbas, 19000, Algeria.

The first conjugate gradient method was proposed by Hestenes and Stiefel (HS) in 1952 [10], to solve a linear system of equations. After the introduction of the nonlinear conjugate gradient method of Fletcher and Reeves (FR) in 1964 [8], many parameters β_k have been proposed which give different conjugate gradient directions d_k . The most famous parameters β_k are those of Polak-Ribiere-Polyak (PRP) [13, 14], Conjugate Descent (CD) [7], Liu-Storey (LS) [11], Dai-Yuan (DY) [3, 4], Hager-Zhang (HZ) [9], Wei et al. (WYL) [18], the MN method proposed by Fan et al. in [6] and Rivaie-Mustafa-Ismail-Leong (RMIL) [15]. Later, many combinations and new families of conjugate gradient methods were proposed, such as those of Sellami and Chaib [16, 17], the different hybridization methods proposed by Mtagulwa and Kaelo [12] and Dalladji et al. [5].

We cite some formulas of the β_k mentioned above:

$$\begin{aligned}\beta_k^{HS} &= \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \quad \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \quad \beta_k^{CD} = -\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}}, \\ \beta_k^{LS} &= -\frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}}, \quad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \quad \beta_k^{RMIL} = \frac{g_k^T y_{k-1}}{\|d_{k-1}\|^2}, \\ \beta_k^{HZ} &= (y_{k-1} - 2d_{k-1} \frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}}) \frac{g_k}{d_{k-1}^T y_{k-1}}, \\ \beta_k^{WYL} &= \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{\|g_{k-1}\|^2}, \quad \beta_k^{MN} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{\mu |g_k^T d_{k-1}| + \|g_{k-1}\|^2} \quad \mu > 1, \\ \beta_k^{SC} &= \frac{(1-\lambda)\|g_k\|^2 + \lambda(-g_k^T d_{k-1})}{-g_{k-1}^T d_{k-1}}, \quad 0 < \lambda < 1, \\ \beta_k^{SC*} &= \frac{\lambda_k \|g_k\|^2 + (1-\lambda_k)g_k^T y_{k-1}}{(1-\lambda_k - \mu_k)\|g_{k-1}\|^2 + (\lambda_k + \mu_k)(y_{k-1}^T d_{k-1})}, \\ &\text{with } 0 < \lambda_k < 1 \text{ and } 0 < \mu_k < 1 - \lambda_k,\end{aligned}$$

where $y_{k-1} = g_k - g_{k-1}$, $s_{k-1} = x_k - x_{k-1}$ and $\|\cdot\|$ denotes the Euclidean vector norm.

The step size α_k is determined using the following strong Wolfe line search conditions

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k g_k^T d_k, \quad (4)$$

$$|g_{k+1}^T d_k| \leq -\sigma g_k^T d_k, \quad (5)$$

where $0 < \rho < \sigma < \frac{1}{2}$.

The aim of this paper is to propose an efficient conjugate gradient method for nonlinear optimization using a new parameter β_k which leads to a new descent direction. The rest of this paper is organized as follows. In Section 2, we give the new formula of β_k and describe the corresponding algorithm. We present a complete analysis of the descent condition of the obtained direction, then, we show the global convergence of the new algorithm using the strong Wolfe line search. Section 3 includes numerical experiments on some examples, considering the well-known test functions in the literature, to show the performance of the considered algorithm. Finally, we end with a conclusion in Section 4.

2. New formula of β_k and description of the corresponding algorithm

The main ingredients of the conjugate gradient method, which play a very important role in convergence analysis and the behavior of the associated algorithm, are the parameter β_k and the displacement step α_k . In this study, motivated by the recent work of Wei et al. [18] and Hager and Zhang [9], we present a new conjugate parameter β_k^{OKB} defined as follows:

$$\beta_k^{OKB} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|d_{k-1}\|} |g_k^T d_{k-1}|}{d_{k-1}^T y_{k-1}}. \quad (6)$$

2.1. Description of the algorithm OKB

Begin algorithm

- Given a starting point $x_1 \in \mathbb{R}^n$ and a parameter $\varepsilon > 0$.
- Set $k = 1$ and compute $d_1 = -g_1$.
- While $\|g_k\| > \varepsilon$ do
 - Find $\alpha_k > 0$ satisfying the strong Wolfe conditions (4) and (5), $\sigma = 0,4$ and $\rho = 10^{-4}$.
 - Take $x_{k+1} = x_k + \alpha_k d_k$.
 - Compute β_{k+1} by the new formula (6).
 - Compute $d_{k+1} = -g_{k+1} + \beta_{k+1} d_k$.
 - Set $k = k + 1$.
- End while.

End algorithm

2.2. Sufficient descent property and global convergence analysis

We start with the following basic assumptions on the objective function in order to establish the global convergence results for the new algorithm.

Assumptions

- (A1) f is bounded below on the level set $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$.
- (A2) In some neighborhood Ω_0 of Ω , f is differentiable and its gradient $g(x)$ is Lipschitz continuous, namely, there exist a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in \Omega_0. \quad (7)$$

Under these assumptions, there exist a constant $c > 0$ such that

$$\|g_k\| \leq c, \forall k \geq 1. \quad (8)$$

Lemma 2.1

Suppose that Assumption (A2) holds, let the sequence $\{x_k\}$ generated by (2) and the step size α_k satisfies strong Wolfe conditions (4) and (5), then

$$g_k^T d_k \leq \frac{2\sigma - 1}{1 - \sigma} \|g_k\|^2. \quad (9)$$

Furthermore, for any k as soon as $g_k \neq 0$, the descent property of d_k is satisfied, i.e.,

$$g_k^T d_k < 0. \quad (10)$$

Proof: The lemma is proved by induction. For $k = 1$, since $d_1 = -g_1$, therefore (9) and (10) are verified.

For some $k > 1$, we suppose that (9) and (10) are true.

Using (3), we get

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_{k+1}^{OKB} g_{k+1}^T d_k, \quad (11)$$

and from (6), we have

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|d_k\|} |g_{k+1}^T d_k|}{d_k^T y_k} g_{k+1}^T d_k. \quad (12)$$

Set $y_k = g_{k+1} - g_k$. Then, from (5) we obtain

$$d_k^T y_k = d_k^T g_{k+1} - d_k^T g_k > \sigma d_k^T g_k - d_k^T g_k = (\sigma - 1) d_k^T g_k > 0. \quad (13)$$

By substituting (5) and (13) into (12), we obtain

$$g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|d_k\|} |g_{k+1}^T d_k|}{(\sigma - 1)d_k^T g_k} (-\sigma d_k^T g_k). \quad (14)$$

We know that

$$0 \leq |g_{k+1}^T d_k| \leq \|g_{k+1}\| \cdot \|d_k\|, \quad (15)$$

this leads to

$$0 \leq \|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|d_k\|} |g_{k+1}^T d_k| \leq \|g_{k+1}\|^2, \quad (16)$$

hence, (14) becomes

$$-\|g_{k+1}\|^2 \leq g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 + \frac{\sigma}{1-\sigma} \|g_{k+1}\|^2. \quad (17)$$

Now, observe that for all $k \geq 1$, we have

$$-\|g_{k+1}\|^2 \leq g_{k+1}^T d_{k+1} \leq \frac{2\sigma - 1}{1 - \sigma} \|g_{k+1}\|^2. \quad (18)$$

Since $0 < \sigma < \frac{1}{2}$, it results that

$$-1 < \frac{2\sigma - 1}{1 - \sigma} < 0. \quad (19)$$

Therefore, $g_{k+1}^T d_{k+1} < 0$. This completes the proof.

Lemma 2.2

Suppose that Assumptions (A1) and (A2) hold. Consider common iterate by (2), with d_k is a sufficient descent direction and α_k is determinate by the strong Wolfe line search condition (4) and (5). Then, the Zoutendjik condition

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty, \quad (20)$$

holds.

Proof: See [19].

Theorem 2.1

Consider the iteration $x_{k+1} = x_k + \alpha_k d_k$, where d_k is defined by (3) and suppose that Assumption (A2) holds, then the new algorithm OKB either stops at stationary point or converges in the sense

$$\lim_{k \rightarrow \infty} \inf \|g_k\| = 0. \quad (21)$$

Proof: We consider for all the demonstration that $\beta_k = \beta_k^{OKB}$.

We suppose that $\lim_{k \rightarrow \infty} \inf \|g_k\| \neq 0$, i.e., for all k , $\exists c > 0$ such that

$$\|g_k\| > c. \quad (22)$$

From (3), we get for all $k \geq 2$

$$\begin{aligned} d_k + g_k &= \beta_k d_{k-1} \\ \|d_k\|^2 + \|g_k\|^2 + 2g_k^T d_k &= \beta_k^2 \|d_{k-1}\|^2 \\ \|d_k\|^2 &= -\|g_k\|^2 - 2g_k^T d_k + \beta_k^2 \|d_{k-1}\|^2. \end{aligned} \quad (23)$$

Dividing by $(g_k^T d_k)^2$, we get

$$\begin{aligned} \frac{\|d_k\|^2}{(g_k^T d_k)^2} &= -\frac{\|g_k\|^2}{(g_k^T d_k)^2} - \frac{2}{g_k^T d_k} + \beta_k^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} \\ &= -\left(\frac{\|g_k\|}{g_k^T d_k} + \frac{1}{\|g_k\|}\right)^2 + \frac{1}{\|g_k\|^2} + \beta_k^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} \\ &\leq \frac{1}{\|g_k\|^2} + \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} \times \frac{\beta_k^2 (g_{k-1}^T d_{k-1})^2}{(g_k^T d_k)^2}. \end{aligned} \quad (24)$$

We remark that $\frac{\beta_k^2 (g_{k-1}^T d_{k-1})^2}{(g_k^T d_k)^2} < 1$.

In fact, we have

$$\beta_k^2 (g_{k-1}^T d_{k-1})^2 - (g_k^T d_k)^2 = (\beta_k g_{k-1}^T d_{k-1} - g_k^T d_k) (\beta_k g_{k-1}^T d_{k-1} + g_k^T d_k),$$

from (13) and (16), we can deduce that β_k is a positive scalar.

From (10) and the positive scalar β_k , we get

$$\beta_k g_{k-1}^T d_{k-1} + g_k^T d_k < 0, \quad (25)$$

and

$$\begin{aligned} \beta_k g_{k-1}^T d_{k-1} - g_k^T d_k &= \beta_k g_{k-1}^T d_{k-1} - g_k^T (-g_k + \beta_k d_{k-1}) \\ &= -\beta_k y_{k-1}^T d_{k-1} + \|g_k\|^2 \\ &= -\|g_k\|^2 + \frac{\|g_k\|}{\|d_{k-1}\|} |g_k d_{k-1}| + \|g_k\|^2 \\ &= \frac{\|g_k\|}{\|d_{k-1}\|} |g_k d_{k-1}| > 0. \end{aligned} \quad (26)$$

Then, from (25) and (26) we get

$$\beta_k^2 (g_{k-1}^T d_{k-1})^2 - (g_k^T d_k)^2 < 0.$$

So, the formula (24) becomes

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{1}{\|g_k\|^2} + \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2},$$

which gives

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} - \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} \leq \frac{1}{\|g_k\|^2},$$

and implies

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} - \frac{\|d_1\|^2}{(g_1^T d_1)^2} \leq \sum_{i=2}^k \frac{1}{\|g_i\|^2}.$$

As, $d_1 = -g_1$, we have

$$\frac{\|d_1\|^2}{(g_1^T d_1)^2} = \frac{1}{\|g_1\|^2},$$

therefore,

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} - \frac{1}{\|g_1\|^2} \leq \sum_{i=2}^k \frac{1}{\|g_i\|^2},$$

which gives

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \sum_{i=1}^k \frac{1}{\|g_i\|^2}, \quad (27)$$

from (22) and (27), we obtain

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} > \frac{c^2}{k}.$$

This implies

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty,$$

thus, contradicting the Zoutendijk condition (20) and guarantying (21), i.e., $\lim_{k \rightarrow \infty} \inf \|g_k\| = 0$. This completes the proof.

3. Numerical experiments

In this section, we present some numerical tests on a set of test functions [1, 2] of unconstrained nonlinear optimization problems using Matlab R2019. Our tests are conducted on DELL PC (Intel (R) Core (TM) i7-7700HQ CPU @ 2.80 GHz, 32 Go RAM), on Windows 10.

The object of these experiments is to show the performance of our new coefficient in comparison with other class of classical existing coefficients. In numerical tests, we consider the **algorithm OKB** based on our new coefficient β_k^{OKB} (6) compared to the Hager-Zhang (HZ) method using β_k^{HZ} [9].

In the table of results we designate by:

- n : the dimension of the problem,
- $iter$: the number of iterations,
- $Time$: the total time in second required to complete the evaluation process.

Test Function	n	β_k^{OKB}		β_k^{HZ}	
		$iter$	$Time$	$iter$	$Time$
DENSCHNA	100	30	0,011245	49	0,019950
	200	30	0,016902	50	0,030366
	500	30	0,032006	50	0,075188
	1000	31	0,062500	51	0,131070
Extended Block Diagonal 1	100	25	0,002885	37	0,004162
	200	25	0,004796	37	0,007083
	500	25	0,034679	38	0,047336
	1000	26	0,038550	38	0,063640
DIXMAANA	100	16	0,006063	19	0,007706
	200	16	0,010358	16	0,012170
	500	17	0,034679	18	0,047336
	1000	17	0,064801	18	0,086842
DIXMAANB	100	14	0,005521	17	0,006647
	200	14	0,009459	20	0,013884
	500	14	0,028822	20	0,050333
	1000	15	0,509552	21	0,098544
Extended Hiebert	100	186	0,021666	782	0,097860
	200	262	0,050333	703	0,142559
	500	150	0,149345	682	0,752066
	1000	302	0,325097	836	1,539170
Extended Rosenbrock	100	127	0,015588	361	0,045206
	200	108	0,031787	700	0,149684
	500	199	0,097230	505	0,487145
	1000	121	0,144109	623	0,781642
DIXMAAND	100	20	0,006672	20	0,008616
	200	21	0,013145	22	0,015209
	500	21	0,048954	22	0,056010
	1000	21	0,082428	21	0,103713
Extended White and Holst	100	69	0,023580	97	0,038436
	200	69	0,044379	107	0,083291
	500	84	0,124510	95	0,173380
	1000	72	0,216234	104	0,373296
Extended Beale	100	109	0,046908	267	0,124510
	200	119	0,092719	247	0,222006
	500	127	0,248663	190	0,437382
	1000	115	0,440869	239	1,063390
Extended Woods	100	219	0,088517	944	0,102763
	200	363	0,057954	1007	0,168555
	500	346	0,133696	1072	0,505063
	1000	286	0,280782	1114	1,169708
Raydan 1	100	12	0,000175	20	0,000353
	200	12	0,000311	20	0,000444
	500	12	0,000438	20	0,000901
	1000	16	0,001061	Failed	Failed

Test Function	n	β_k^{OKB}		β_k^{HZ}	
		$iter$	$Time$	$iter$	$Time$
Extended Powell	100	433	0,208014	Failed	Failed
	200	402	0,364384	Failed	Failed
	500	497	1,099740	Failed	Failed
	1000	615	2,729780	Failed	Failed
DIXMAANC	100	16	0,007667	19	0,008081
	200	17	0,012793	18	0,014042
	500	18	0,035370	19	0,049259
	1000	18	0,074232	20	0,095920
Power	100	1140	0,064120	1814	0,138626
	200	2545	0,201966	3487	0,369626
	500	6544	0,830418	9740	1,672919
	1000	10804	2,824616	18669	5,641824
Quadratic Diagonal Perturbed	100	159	0,016168	220	0,028212
	200	200	0,052648	426	0,079868
	500	292	0,026646	384	0,067986
	1000	336	0,081887	548	0,154064
Rodenstein and Roth	100	82	0,016210	169	0,037357
	200	82	0,092719	265	0,222006
	500	69	0,055907	187	0,238901
	1000	96	0,176877	118	0,290437
HARKERP	100	299	0,017223	510	0,029777
	200	383	0,057954	952	0,168555
	500	141	0,027043	243	0,098434
	1000	219	0,059078	362	0,165848
Extended Tridiagonal 1	100	54	0,019141	72	0,029115
	200	54	0,033980	Failed	Failed
	500	54	0,083032	Failed	Failed
	1000	65	0,198614	Failed	Failed

3.1. Commentaries

From the results obtained in the tables above, it is clear that our new algorithm based on the β_k^{OKB} parameter is more efficient than the HZ method in terms of number of iterations and computation time. There is a significant reduction in the number of iterations using the OKB algorithm compared to HZ algorithm. Furthermore, our algorithm is more competitive in terms of computation time with the HZ method. On the other hand, when the size of some examples becomes large, the HZ algorithm fails to provide the optimal solution after a number of iterations $k_{\max} = 50000$.

4. Conclusion

We have proposed a new β_k , also we have provided proof of the global convergence of our proposed algorithm OKB for nonlinear functions using the strong Wolfe line search. The numerical tests carried out confirm the effectiveness of the new algorithm OKB. It has a good performance compared to Hager-Zhang (HZ) method for the selected list of test functions.

REFERENCES

1. N. Andrei, *Nonlinear conjugate gradient methods for unconstrained optimization*, Springer Optimization and its Applications, vol. 158, 2020.
2. N. Andrei, *An unconstrained optimization test functions collection*, Adv. Model. Optim., vol. 10, pp. 147–161, 2008.
3. Y.H. Dai, Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J. Optim., vol. 10, no. 1, pp. 177–182, 1999.
4. Y.H. Dai, Y. Yuan, *An efficient hybrid conjugate gradient method for unconstrained optimization*, Ann. Oper. Res., vol. 103, pp. 33–47, 2001.
5. S. Delladji, M. Belloufi, B. Sellami, *New hybrid conjugate gradient method as a convex combination of FR and BA methods*, Journal of Information and Optimization Sciences, vol. 42, no. 3, pp. 591–602, 2021.
6. H. Fan, Z. Zhu, A. Zhou, *A new descent nonlinear conjugate gradient method for unconstrained optimization*, Applied Mathematics, Vol. 2, no. 9, pp. 1119–1123, 2011.
7. R. Fletcher, *Practical methods of optimization. Unconstrained Optimization*, vol. 1, Wiley, New York, 1987.
8. R. Fletcher, C. M. Reeves, *Function minimization by conjugate gradients*, Comput. J., vol. 7, no. 2, pp. 149–154, 1964.
9. W.W. Hager, H. Zhang, *A survey of nonlinear conjugate gradient methods*, Pacific journal of Optimization, vol. 2, pp. 35–58, 2006.
10. M. R. Hestenes, E. Steifel, *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Bur. Stand., vol. 49, no. 6, pp. 409–436, 1952.
11. Y. Liu, C. Storey, *Efficient generalized conjugate gradient algorithms, part 1: theory*, J. Optim. Theory Appl, vol. 69, no. 1, pp. 129–137, 1991.
12. P. Mtagulwa, P. Kaelo, *A convergent modified HS-DY hybrid conjugate gradient method for unconstrained optimization problems*, Journal of Information and Optimization Sciences, vol. 40, no. 1, pp. 97–113, 2019.
13. E. Polak, G. Ribiere, *Note sur la convergence des méthodes de directions conjuguées*, Rev. Française Informatique Recherche Opérationnelle, vol. 16, pp. 35–43, 1969.
14. B. T. Polyak, *The conjugate gradient method in extreme problems. U.S.S.R*, Comput. Math. Phys., vol. 9, pp. 94–112, 1969.
15. M. Rivaie, M. Mustafa, W. J. Leong, M. Ismail, *A new class of nonlinear conjugate gradient coefficients with global convergence properties*, Applied Mathematics and Computation, vol. 218, no. 22, pp. 11323–11332, 2012.
16. B. Sellami, Y. Chaib, *A new family of globally convergent conjugate gradient methods*, Ann. Oper. Res. Springer., vol. 241, pp. 497–513, 2016.
17. B. Sellami, Y. Chaib, *New conjugate gradient method for unconstrained optimization*, RAIRO Operations Research, vol. 50, pp. 1013–1026, 2016.
18. Z. Wei, S. Yao, L. Liu, *The convergence properties of some new conjugate gradient methods*, Applied Mathematics and Computation 183 1341–1350, 2006.
19. G. Zoutendijk, *Nonlinear programming, computational methods*. In: Abadie. J. (ed.), Integer and Nonlinear Programming, 1970.