Nonlinear Dynamics and Systems Theory, **25** (1) (2025) 91–100

Efficient Descent Direction of a Conjugate Gradient Algorithm for Nonlinear Optimization

Mohamed Lamine Ouaoua* and Samia Khelladi

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

Received: February 21, 2024; Revised: January 9, 2025

Abstract: We propose an efficient variant of the conjugate gradient method for nonlinear optimization based on a new parameter β_k. We show that the new search direction fills the sufficient descent condition and we prove the global convergence of the corresponding algorithm using the strong Wolfe inexact line search. The established numerical results show that the new algorithm is more efficient in comparison with the standard Fletcher-Reeves method in terms of either the iteration number or CPU time.

Keywords: *unconstrained optimization; conjugate gradient method; descent direction; inexact line search; global convergence.*

Mathematics Subject Classification (2020): 65K05, 90C26, 90C30, 93A30.

1 Introduction

Consider the following unconstrained nonlinear optimization problem:

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

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

Conjugate gradient methods are efficient to solve unconstrained optimization problem (1), especially for large scale problems. These methods generate the following sequence:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

* Corresponding author: mailto:mohamedlamine.ouaoua@univ-setif.dz

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

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

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

Different conjugate gradient methods correspond to different values of the coefficient β_k, a survey of these methods was given by Hager an Zhang in [11].

Recently, a great contribution in the area of conjugate gradient methods and their application has been done by Andrei in [1].

Among the well known formulas of β_k, we can cite Hestenes-Stiefel (HS) [13], Fletcher-Reeves (FR) [10], Polak-Ribière-Polyak (PRP) [15,16], Conjugate Descent-Fletcher (CD) [9], Liu-Story (LS) [14] and Dai-Yuan (DY) [4], which are given as follows:

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}},$$

$$\beta_k^{FR} = \frac{\| g_k \|^2}{\| g_{k-1} \|^2},$$

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\| g_{k-1} \|^2},$$

$$\beta_k^{CD} = -\frac{\| g_k \|^2}{g_{k-1}^T d_{k-1}},$$

$$\beta_k^{LS} = -\frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}},$$

$$\beta_k^{DY} = \frac{\| g_{k-1} \|^2}{d_{k-1}^T y_{k-1}},$$

where $y_{k-1} = g_k - g_{k-1}$.

If f is a strongly convex quadratic function and the line search is exact, then all the above formulas of β_k are the same.

In the general case, where f is non-quadratic, the algorithms corresponding to each β_k have different numerical performances.

There are many other conjugate gradient methods such as those where the scalar is given by parametric formulas like the β_k proposed by Sellami and Chaib in [17,18].

Several researchers have also proposed hybrid conjugate gradient methods combining the existing β_k [3,5–8,12] .

There are many convergence results with some line search conditions which have been widely studied, there the method can guarantee the descent property of each direction which provided the step length α computed by carrying out a line search and it satisfies the strong Wolfe conditions such that

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha g_k^T d_k, \tag{4}$$

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

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

The aim of this paper is to ameliorate the conjugate gradient method for nonlinear optimization using a new parameter which leads to a new descent direction.

The rest of the paper is organized as follows. In Section 2, we define the formula of the new conjugate gradient coefficient using the Fletcher-Reeves formula with a modification in its denominator, then we give the description of the corresponding algorithm. In the second part, we present a complete analysis of the descent condition of the obtained direction, then we show the global convergence of the corresponding algorithm. Section 3 contains numerical experiments on some examples considering the well known test functions in the literature. Finally, we end with a conclusion in Section 4.

2 Convergence Analysis of the Algorithm Based on the New Parameter β_k

In this section, we propose a new conjugate gradient coefficient using the Fletcher-Reeves formula [10] with a modification in its denominator. This coefficient is defined as follows:

$$\beta_k^{MSD} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2 + \mu \left| g_k^T d_{k-1} \right|}, \tag{6}$$

where $\mu > 0$.

Recall that the case $\mu = 0$ corresponds to the Fletcher-Reeves coefficient [10].

2.1 Description of the conjugate gradient algorithm

2.1.1 Algorithm MSD

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 $(4, 5)$.
- Take $x_{k+1} = x_k + \alpha_k d_k$.
- Compute β_{k+1} by the new formula (6).
- Set $d_{k+1} = -g_{k+1} + \beta_{k+1} d_k$ and $k = k + 1$.
- End while.

End algorithm.

2.2 Sufficient descent property and global convergence analysis

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

2.2.1 Assumptions

- (i) f is a lower bounded function on the level set

$$\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}.$$

- (ii) In some neighborhood Ω_0 of Ω, f is differentiable and its gradient $g(x)$ is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

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

Under these assumptions, there exists a constant $\varepsilon > 0$ such that

$$\|g_k\| \leq \varepsilon, \forall k. \tag{8}$$

To prove the global convergence of the nonlinear conjugate gradient methods, we use the following lemma, called the Zoutendijk condition [19].

Lemma 2.1 *[19] Suppose that the assumptions $(i), (ii)$ hold. Let the sequence $\{x_k\}$ be generated by (2) and d_k satisfy $g_k^T d_k < 0$. If α_k is determined by the Wolfe line search conditions, then we have*

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

Lemma 2.2 *Suppose that the assumption (ii) holds, let the sequence $\{x_k\}$ be generated by (2) and the step length α_k satisfy the strong Wolfe conditions with $0 < \sigma < \frac{1}{2}$, then for any k,*

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

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

$$g_k^T d_k < 0. \tag{10}$$

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

For some $k > 1$, we suppose that (9) and (10) are true. By using (3), we get

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = -1 + \frac{\beta_{k+1} g_{k+1}^T d_k}{\|g_{k+1}\|^2}. \tag{11}$$

From the second strong Wolfe condition (5) and (11), we get

$$-1 + \sigma \frac{\beta_{k+1} g_k^T d_k}{\|g_{k+1}\|^2} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 - \sigma \frac{\beta_{k+1} g_k^T d_k}{\|g_{k+1}\|^2}. \tag{12}$$

By using (6), we have

$$-1 + \sigma \frac{g_k^T d_k}{\|g_k\|^2 + \mu \left|g_{k+1}^T d_k\right|} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 - \sigma \frac{g_k^T d_k}{\|g_k\|^2 + \mu \left|g_{k+1}^T d_k\right|}. \tag{13}$$

Observe that for all k, we have

$$\frac{1}{\|g_k\|^2 + \mu \left|g_{k+1}^T d_k\right|} \leq \frac{1}{\|g_k\|^2}. \tag{14}$$

By introducing (14) in (13), we get

$$-1 + \sigma \frac{g_k^T d_k}{\|g_k\|^2} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 - \sigma \frac{g_k^T d_k}{\|g_k\|^2}. \tag{15}$$

From (9), it follows that for all k,

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

which implies

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

This gives the formula (9).

Since $0 < \sigma < \frac{1}{2}$, it results from (9) that $g_k^T d_k < 0$.

This completes the proof of the lemma.

Theorem 2.1 *Consider the sequence $\{x_k\}$ generated by (2), where d_k is defined by (3) and satisfies $g_k^T d_k < 0$ and suppose that the assumptions (i) and (ii) hold. Then the Algorithm MSD either stops at a stationary point or converges in the sense that*

$$\lim_{k \longrightarrow \infty} \inf \|g_k\| = 0. \tag{16}$$

Proof. From the second strong Wolfe condition (5) and (9), we get

$$\left|g_k^T d_{k-1}\right| \leq -\sigma g_{k-1}^T d_{k-1} \leq \frac{\sigma}{1-\sigma} \|g_{k-1}\|^2, \tag{17}$$

using (3), (14) and (17), we obtain

$$\begin{aligned} \|d_k\|^2 &= \|g_k\|^2 - 2\beta_k g_k^T d_{k-1} + \beta_k^2 \|d_{k-1}\|^2 \\ &\leq \|g_k\|^2 + \beta_k \frac{2\sigma}{1-\sigma} \|g_{k-1}\|^2 + \beta_k^2 \|d_{k-1}\|^2 \\ &\leq \|g_k\|^2 + \frac{2\sigma}{1-\sigma} \|g_k\|^2 + \beta_k^2 \|d_{k-1}\|^2 \\ &\leq \left(\frac{1+\sigma}{1-\sigma}\right) \|g_k\|^2 + \beta_k^2 \|d_{k-1}\|^2 . \end{aligned} \tag{18}$$

After defining $\tau = \frac{1+\sigma}{1-\sigma}$ and applying (18) repeatedly, and using (14), it follows that

$$\begin{aligned} \|d_k\|^2 &\leq \tau\|g_k\|^2 + \beta_k^2 \|d_{k-1}\|^2 \\ &\leq \tau \left[\|g_k\|^2 + \beta_k^2 \|g_{k-1}\|^2 + \beta_k^2\beta_{k-1}^2 \|g_{k-2}\|^2 + \cdots + \beta_k^2\beta_{k-1}^2 \cdots \beta_3^2 \|g_2\|^2\right] \\ &\quad +\beta_k^2\beta_{k-1}^2 \cdots \beta_2^2 \|d_1\|^2 \\ &\leq \tau \left[\|g_k\|^2 + \frac{\|g_k\|^4}{\|g_{k-1}\|^2} + \frac{\|g_k\|^4}{\|g_{k-2}\|^2} + \cdots + \frac{\|g_k\|^4}{\|g_2\|^2}\right] + \frac{\|g_k\|^4}{\|g_1\|^2} \\ &\leq \tau\|g_k\|^4 \left[\sum_{i=2}^{k} \frac{1}{\|g_i\|^2}\right] + \frac{\|g_k\|^4}{\|g_1\|^2} \\ &\leq \tau\|g_k\|^4 \left[\sum_{i=1}^{k} \frac{1}{\|g_i\|^2}\right]. \end{aligned} \tag{19}$$

Now, if (16) is not true, then there exists a constant $\varepsilon > 0$ such that $\|g_k\| > \varepsilon$ for all k. From (19), we obtain

$$\frac{\|g_k\|^4}{\|d_k\|^2} \geq \frac{\varepsilon^2}{\tau k},$$

which implies

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} = \infty. \tag{20}$$

By using (9), we have

$$\frac{(g_k d_k)^2}{\|d_k\|^2} \geq \left(\frac{1-2\sigma}{1-\sigma}\right)^2 \frac{\|g_k\|^4}{\|d_k\|^2}, \tag{21}$$

which implies

$$\sum_{k=1}^{\infty} \frac{(g_k d_k)^2}{\|d_k\|^2} = \infty, \tag{22}$$

thus contradicting the Zoutendijk condition.

This completes the proof of the theorem.

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 language.

The objective of these experiments is to show the performance of our new coefficient in comparison with other class of existing classical coefficients. In numerical tests, we consider the *Algorithm MSD* based on our new coefficient β_k^{MSD} (6) compared with the FR method with β_k^{FR} [10].

In the tables of results, we designate by

- n: the size of the problem

- *iter*: the number of iterations

- *time*: the total time in seconds required to complete the evaluation process.

Test function	Size	β_k^{MSD}		β_k^{FR}	
	n	*iter*	*time*	*iter*	*time*
Raydan 1	100	8	0,000411	15	0,000535
	200	8	0,000604	15	0,000852
	500	8	0,001085	15	0,001769
	1000	8	0,002022	15	0,003233
Raydan 2	100	16	0,000421	52	0,001469
	200	16	0,000543	57	0,002259
	500	21	0,001113	3241	0,232550
	1000	21	0,001937	NAN	NAN
Diagonal 4	100	41	0,003746	63	0,005513
	200	42	0,007477	60	0,009580
	500	44	0,016702	62	0,020083
	1000	44	0,038645	99	0,081462
Extended Woods	100	369	0,062696	839	0,162356
	200	471	0,131264	782	0,233382
	500	430	0,271904	659	0,472802
	1000	430	0,029804	932	1,597367
HIMMELBC	100	30	0,004411	42	0,004066
	200	30	0,006783	43	0,007018
	500	32	0,014692	43	0,016906
	1000	32	0,054439	44	0,040352
DIXMAANC	100	14	0,005565	32	0,011920
	200	15	0,012137	23	0,016665
	500	18	0,030703	21	0,033251
	1000	16	0,054192	17	0,073482
HARKERP	100	91	0,004411	302	0,014902
	200	98	0,006783	102	0,007951
	500	108	0,014692	156	0,020768
	1000	255	0,054439	NAN	NAN
PROD1	100	26	0,020014	36	0,022772
	200	26	0,075946	37	0,099651
	500	13	0,262461	13	0,274593
	1000	13	1,105908	14	1,160563

Test function	Size	β_k^{MSD}		β_k^{FR}	
	n	*iter*	*time*	*iter*	*time*
Extended Block Diagonal 1	100	23	0,002571	103	0,016227
	200	24	0,004266	107	0,026036
	500	24	0,013149	103	0,064142
	1000	25	0,028290	114	0,144926
Extended Maratos	100	41	0,002943	42	0,005326
	200	41	0,005166	42	0,009220
	500	41	0,012851	42	0,021821
	1000	42	0,029804	43	0,045364
DIXMAANB	100	13	0,005753	19	0,007854
	200	13	0,011356	20	0,014953
	500	12	0,023140	20	0,031973
	1000	12	0,041619	22	0,066770
Extended Beale	100	66	0,033441	79	0,041068
	200	66/	0,061550	128	0,136998
	500	70	0,155799	90	0,220902
	1000	70	0,312607	132	0,665937
Extended White and holst	100	48	0,021397	63	0,025315
	200	48	0,037467	74	0,054997
	500	51	0,093806	144	0,293700
	1000	52	0,187531	78	0,279982
Quadratic Diagonal Perturbed 1	100	102	0,005432	126	0,008635
	200	155	0,009727	1073	0,065435
	500	280	0,029289	1332	0,145182
	1000	421	0,078164	1983	0,411533
DIXMAANA	100	12	0,005473	19	0,007968
	200	12	0,010762	21	0,015383
	500	13	0,025386	20	0,031282
	1000	13	0,046094	20	0,064289
DENSCHNA	100	35	0,009824	58	0,016107
	200	35	0,020065	61	0,031975
	500	35	0,041322	59	0,066094
	1000	36	0,081710	64	0,149320
Freudenstein and Roth	100	136	0,018940	265	0,038783
	200	115	0,026829	171	0,038423
	500	119	0,065969	145	0,085392
	1000	111	0,142784	112	0,143777
Nondiag	100	138	0,014461	NAN	NAN
	200	136	0,011197	NAN	NAN
	500	240	0,043510	NAN	NAN
	1000	835	0,243155	NAN	NAN

Remark 3.1 In the results tables, NAN means that the algorithm does not display the optimal solution after a maximum number of iterations k_{max}.

3.1 Commentaries

From the results obtained in the tables above, it is clear that our new algorithm based on the β_k^{MSD} parameter is more efficient than the FR method in terms of the number of iterations and computation time. There is a significant reduction in the number of iterations when using the algorithm MSD compared to FR. On the other hand, when the size of certain examples becomes large $n > 1000$, the FR algorithms fail to provide the optimal solution after the number of iterations $k_{max} = 50000$.

4 Conclusion

We proposed a new β_k, and also provided the proof of the global convergence of the new *Algorithm MSD.* We have proven the effectiveness of this algorithm based on the new β_k^{MSD}. It has a good performance compared with other conjugate gradient methods such as the FR method when considering the selected list of test problems.

References

[1] M. Andrei. *Nonlinear Conjugate Gradient Methods for Unconstrained Optimization.* Springer Optimization and its Applications, Vol. 158, 2020.

[2] M. Andrei. An unconstrained optimization test functions collection. *Adv. Model. Optim.* **10** (2008) 1–20.

[3] S. Ben Hanachi, B. Sellami and M. Belloufi. New iterative conjugate gradient method for nonlinear unconstrained optimization. *RAIRO-Oper.* **56** (2022) 2315–2327.

[4] Y. H. Dai and Y. Yuan. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J. Optim.* **10** (1) (1999) 177–182.

[5] Y. H. Dai and Y. Yuan. An efficient hybrid conjugate gradient method for unconstrained optimization. *Ann. Oper. Res.* **103** (2001) 33–47.

[6] S. Delladji, M. Belloufi and B. Sellami. New hybrid conjugate gradient method as a convex combination of FR and BA methods. *Journal of Information and Optimization Sciences.* **42** (3) (2021) 591–602.

[7] S. Djordjevic. New hybrid conjugate gradient method as a convex combination of LS and FR methods. *Acta Mathematica Scientia.* **39B** (1) (2019) 214–378.

[8] S. Djordjevic. New hybrid conjugate gradient method as a convex combination of HS and FR methods. *Journal of Applied Mathematics and Computation.* **2** (9) (2019) 366–378.

[9] R. Fletcher. *Practical Methods of Optimization.* Unconstrained optimization. Vol. 1, Wiley, New York, 1987.

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

[11] W. W. Hager and H. Zhang. A survey of nonlinear conjugate gradient methods. *Pacific Journal of Optimization* **2** (2006) 35–58.

[12] A. Hallal, M. Belloufi and B. Sellami. An efficient new hybrid CG-method as convex combination of DY and CD and HS algorithms. *RAIRO-Oper. Res.* **56** (2022) 4047–4056.

[13] M. R. Hestenes and E. Steifel. Methods of conjugate gradient for solving linear systems. *Res. Natl. Bur. Stand.* **49**(6) (1952) 409–436.

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

[15] E. Polak and G. Ribiere. Note sur la convergence des méthodes de directions conjuguées. *Rev. Française d'Informatique et de Recherche Opérationnelle.* **16** (2) (1969) 35–43.

[16] B. T. Polyak. The conjugate gradient method in extrem problems. *U.S.S.R. Comput. Math. Phys.* **9** (1969) 94–112.

[17] B. Sellami and Y. Chaib. A new family of globally convergent conjugate gradient methods. *Ann. Oper. Res. Springer.* **241** (2016) 497–513.

[18] B. Sellami and Y. Chaib. New conjugate gradient method for unconstrained optimization. *RAIRO Operations Research.* **50** (2016) 1013–1026.

[19] G. Zoutendijk. Nonlinear programming, computational methods. In: Integer and Nonlinear Programming (J. Abadie (ed.), 1970, 37–86.