Journal of King Saud University – Science 34 (2022) 102281

Contents lists available at ScienceDirect

Journal of King Saud University – Science

journal homepage: www.sciencedirect.com

Original article

A criterion for the global convergence of conjugate gradient methods under strong Wolfe line search

Osman O.O. Yousif [a], Mogtaba A.Y. Mohammed [b,c,*], Mohammed A. Saleh [d], Murtada K. Elbashir [e]

[a] Department of Mathematics, Faculty of Mathematical and Computer Sciences, University of Gezira, Sudan
[b] Department of Mathematics, Faculty of Sciences AlZulfi, Majmaah University, Majmaah 11952, Saudi Arabia
[c] Department of Mathematics and Applied Mathematics University of Pretoria, Pretoria 0002, South Africa
[d] Department of Computer, College of Science and Arts, Qassim University, Ar Rass, Saudi Arabia
[e] Department of Information Systems, College of Computer and Information Sciences, Jouf University, Sakaka 72441, Kingdom of Saudi Arabia

ARTICLE INFO

Article history:
Received 25 April 2021
Revised 24 June 2022
Accepted 12 August 2022
Available online 24 August 2022

Keywords:
Unconstrained optimization problems
Conjugate gradient methods
Strong Wolfe line search
Sufficient descent property
Global convergence

ABSTRACT

From 1952 until now, the sufficient descent property and the global convergence of conjugate gradient (CG) methods have been studied extensively. However, the sufficient descent property and the global convergence of some CG methods such as the method of Polak, Ribière, and Polyak (PRP) and the method of Hestenes and Stiefel (HS) have not been established yet under the strong Wolfe line search. In this paper, based on Yousif (Yousif, 2020) we present a criterion that guarantees the generation of descent search directions property and the global convergence of CG methods when they are applied under the strong Wolfe line search. Moreover, the PRP and the HS methods are restricted in order to satisfy the presented criterion, so new modified versions of PRP and HS are proposed. Finally, to support the theoretical proofs, a numerical experiment is done.

1. Introduction

Unconstrained optimization problems usually arise in various fields of science, engineering, and economics. They are mathematically formulated as

$$\min_{x \in R^n} f(x), \tag{1.1}$$

where $f : R^n \to R$ is a continuously differentiable function. The methods of the conjugate gradient are widely used to solve problems (1.1), this is due to their simplicity and small footprint. It should be mentioned that optimization problems as in (Eq. (1.1)) are also solved using non-gradient methods especially successfully using different population based heuristics. The previous methods use the following iterative expression:

$$x_{k+1} = x_k + \alpha_k d_k, k = 0, 1, 2, \cdots, \tag{1.2}$$

Such that α_k represents the step length that is a CG method takes in each step toward the search direction d_k. Strong Wolfe line search is one of the most used methods in practical computations for computing α_k, in which α_k satisfies

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k \tag{1.3}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \le \sigma |g_k^T d_k| \tag{1.4}$$

Such that g_k represents the gradient of the nonlinear function f at the value x_k and $0 < \delta < \sigma < 1$ and d_k is the search direction given by:

$$d_k = \begin{cases} -g_k, if k = 0, \\ -g_k + \beta_k d_{k-1}, if\ k \geqslant 1, \end{cases} \tag{1.5}$$

Such that β_k is the factor that determines how the conjugate gradient methods differ. Some of the very well-known formulas attributed to Hestenes-Stiefel (HS) (Hestenes and Stiefel, 1952), Fletcher-Reeves (FR) (Fletcher and Reeves, 1964) and Polak-Ribière-Polyak (PRP) (Polyak, 1969; Polak and Ribière, 1969). These formulas are given by

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})}$$

* Corresponding author.
E-mail addresses: osman_om@hotmail.com (O.O.O. Yousif), mogtaba.m@mu.edu.sa (M.A.Y. Mohammed), m.saleh@qu.edu.sa (M.A. Saleh), mkelfaki@ju.edu.sa (M.K. Elbashir).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksus.2022.102281
1018-3647/

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

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

respectively. Other formulas are conjugate descent (CD) (Fletcher, 1987), Liu-Storey (LS) (Liu and Storey, 1992), and Dai-Yuan (DY) (Dai and Yuan, 2000). For more formulas for the coefficient β_k see (Abubakar et al., 2022; Yuan and Lu, 2009; Zhang, 2009; Rivaie et al., 2012; Hager and Zhang, 2005; Dai, 2002; Yuan and Sun, 1999; Salleh et al., 2022; Dai, 2016; Wei et al., 2006, Wei et al., 2006).

To guarantee that every search direction generated by a CG method is descent, the sufficient descent property

$$g_k^T d_k \le -C\|g_k\|^2, \forall k \ge 0 \text{ and a constant } C > 0, \tag{1.6}$$

is needed.

The global convergence and descent directions property of the FR method are established using both exact (Zoutendijk and Abadie, 1970) and strong Wolfe line search (Al-Baali, 1985) on general functions. The PRP and the HS methods with exact line search can cycle infinitely without approaching a solution which implies that they both do not have global convergence for general functions (Powell, 1984). Nevertheless, the good performance of the PRP and the HS in practice, that is, due to self-restarting property, both methods are preferred to the FR method. To establish the convergence of them with the strong Wolfe line search, Powell (Powell, 1986) suggested restricting them to be non-negative. Motivated by Powell's suggestion (Powell, 1986), Gilbert and Nocedal (Gilbert and Nocedal, 1992) conducted an elegant analysis and established that they are globally convergent if they are restricted to be non-negative and the step length satisfies the sufficient descent condition. Further studies on global convergence properties of CG methods are of Hu and Storey (Hu and Storey, 1991), Liu et al (Zoutendijk and Abadie, 1970), and Touati-Ahmed and Storey (Touati-Ahmed and Storey, 1990) among others.

Recently, Yousif (Yousif, 2020) gave detailed proof for the sufficient descent property and the global convergence of the modified method of Rivaie; Mamat, Ismail, and Leong (RMIL +) (Rivaie et al., 2012). In this author's work, the coefficient is given by

$$\beta_k^{RMIL+} = \begin{cases} \frac{g_k^T(g_k - g_{k-1})}{\|d_{k-1}\|^2}, if 0 \le g_k^T g_{k-1} \le \|g_k\|^2, \\ 0, otherwise. \end{cases} \tag{1.7}$$

The proof is based on the inequality

$$\frac{\|g_k\|}{\|d_k\|} < 2, k \ge 0, \tag{1.8}$$

In the above setting the RMIL + method generated $\{g_k\}$ and $\{d_k\}$ under the application of strong Wolfe line search in the case of $\sigma \in [0, \frac{1}{4}]$.

In this paper, inspired by Yousif (Yousif, 2020), we present a criterion that guarantees the descent property and the global convergence of each CG method satisfying this criterion. This is presented in Sections 2. In Section 3, based on this criterion, we propose modified versions of PRP and HS methods. Finally, in Section 4, to show the efficiency of the proposed modified methods in practical computation, they are compared with PRP, HS, FR, and RMIL + methods.

2. A new criterion guarantees sufficient descent and global convergence

In this section, we firstly show that for every CG method whose coefficient β_k satisfies

$$|\beta_k| \le \mu \frac{\|g_k\|^2}{\|d_{k-1}\|^2}, \text{for} k \ge 1 \text{and a real number} \mu \ge 1, \tag{2.1}$$

the inequality (1.8) holds true. Secondly, we prove the sufficient descent property and the global convergence of any CG method whose coefficient β_k satisfies (2.1) under the application of strong Wolfe line search in the case of $\sigma \in [0, \frac{1}{4\mu}]$.

2.1. The sufficient descent property

Before we prove the desired property, we first note that for every-two positive real numbers σ and $\mu \ge 1$, we have

$$0 < \sigma < \frac{1}{4\mu} \Rightarrow -2 < 2(2\mu\sigma - 1) < -1$$

$$\Rightarrow -1 < 2\mu\sigma - 1 < \frac{-1}{2}$$

$$\Rightarrow \frac{1}{2} < 1 - 2\mu\sigma < 1$$

$$\Rightarrow 1 < \frac{1}{1 - 2\mu\sigma} < 2 \tag{2.2}$$

Theorem 2.1: Assume that $\{g_k\}$ *and* $\{d_k\}$ are generated by a CG method such that β_k satisfies (2.1) under the application of strong Wolfe line search in the case of $\sigma \in [0, \frac{1}{4\mu}]$. Then (1.8) holds.

Proof: We follow the induction argument. For $k = 0$, (1.5) shows that (1.8) is satisfied. Now , suppose that (1.8) is true for $k \ge 1$, rewrite equation (1.5) for $k + 1$ and multiply the resulting equation by g_{k+1}^T, we get

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

Applying the triangle inequality, we get

$$\|g_{k+1}\|^2 \le |g_{k+1}^T d_{k+1}| + |\beta_{k+1} g_{k+1}^T d_k|$$

Using the condition (1.4), we obtain

$$\|g_{k+1}\|^2 \le |g_{k+1}^T d_{k+1}| + \sigma|\beta_{k+1}||g_k^T d_k|$$

Substitute (2.1) for β_{k+1} and use C—S inequality, we get

$$\|g_{k+1}\|^2 \le \|g_{k+1}\|\|d_{k+1}\| + \mu\sigma\|g_{k+1}\|^2 \frac{\|g_k\|}{\|d_k\|}. \tag{2.3}$$

Dividing both sides of (2.3) by $\|g_{k+1}\|$ and then applying the induction hypothesis (1.8), we come to

$$\|g_{k+1}\| < \|d_{k+1}\| + 2\mu\sigma\|g_{k+1}\|$$

which leads to

$$\|g_{k+1}\|(1 - 2\mu\sigma) < \|d_{k+1}\|$$

Since $(1 - 2\mu\sigma) > 0$ and $\frac{1}{1-2\mu\sigma} < 2$ (see (2.2)), we come to

$$\frac{\|g_{k+1}\|}{\|d_{k+1}\|} < \frac{1}{1 - 2\mu\sigma} < 2$$

thus, the proof is complete.

Now, we are able to establish the sufficient descent property (1.6) under the condition (2.1). This is the topic of the following theorem

Theorem 2.2: *Assume that* $\{g_k\}$ *and* $\{d_k\}$ *are generated by a CG method such that* β_k *satisfies (2.1) under the application of strong Wolfe line search in the case of* $\sigma \in [0, \frac{1}{4\mu}]$. *Then the sufficient descent property (1.6) holds true.*

Proof: For $k = 0$, the result is clear by using (1.5). Consider the case $k > 0$.

From (1.5), we have

$$g_k^T d_k = -\|g_k\|^2 + \beta_k g_k^T d_{k-1}$$

$$\le -\|g_k\|^2 + |\beta_k||g_k^T d_{k-1}|$$

Applying the strong Wolfe condition (1.4), we get

$$g_k^T d_k \le -\|g_k\|^2 + \sigma|\beta_k||g_{k-1}^T d_{k-1}|.$$

Using Cauchy-Schwartz inequality $|g_{k-1}^T d_{k-1}| \le \|g_{k-1}\|\|d_{k-1}\|$ and then substituting (2.1) and (1.8), we come to

$$g_k^T d_k \le -\|g_k\|^2 + \mu\sigma\|g_k\|^2 \frac{\|g_{k-1}\|}{\|d_{k-1}\|}$$

$$< (-1 + 2\mu\sigma)\|g_k\|^2,$$

which means

$$g_k^T d_k < -C\|g_k\|^2, \tag{2.4}$$

where $C = 1 - 2\mu\sigma > 0$ (see (2.2)). Thus the proof is complete.

2.2. The global convergence

Now, based on the following assumption on the objective function f, we establish the global convergence under strong Wolfe line search with $0 < \sigma < \frac{1}{4\mu}$ of every CG method whose coefficient satisfies (2.1).

Assumption 2.1

(i) Define $\Omega = \{x \in R^n : f(x) \le f(x_0)\}$ and assume that Ω is bounded for all initial points x_0.
(ii) Let N be a neighborhood of Ω and assume that $f \in C(\mathrm{N})$ such that for some $l > 0$
$\|g(x)-g(y)\| \le l\ \|x-y\|,\ \forall\ x,y \in N$.

Under this assumption, Zoutendijk (Zoutendijk and Abadie, 1970) proved the following results.

Lemma 2.1 Let Assumption 2.1 is given. The for any conjugate gradient method in the forms (1.2)-(1.5) such that α_k is computed according to strong Wolfe line search. Then

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{2.5}$$

From (2.4), we get

$C^2\|g_k\|^4 < (g_k^T d_k)^2$, for all $k \ge 0$

which leads to

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \frac{1}{C^2} \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2}. \tag{2.6}$$

From (2.5) and (2.6) together, we come to.

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{2.7}$$

Therefore, based on Assumption 2.1, we deduce that if the sequences $\{g_k\}$ and $\{d_k\}$ are generated by a CG method with coefficient β_k satisfying (2.1) when it is applied under the strong Wolfe line search with $0 < \sigma < \frac{1}{4\mu}$, then (2.7) holds.

The following lemma will be used in the proof of the global convergence

Lemma 2.2: *Suppose that* $\{g_k\}$ *and* $\{d_k\}$ *are generated by any CG method such that* β_k *satisfies (2.1) under the application of the strong Wolfe line search with* $0 < \sigma < \frac{1}{4\mu}$. *Then there exists a positive constant* $C_1 > 1$ *such that*

$$g_k^T d_k \ge -C_1\|g_k\|^2 \tag{2.8}$$

Proof: Multiplying (1.5) by g_k^T and then applying the triangle inequality, we obtain

$$|g_k^T d_k| \le \|g_k\|^2 + |\beta_k||g_k^T d_{k-1}|$$

Substituting (2.1) and applying the strong Wolfe condition (1.4) and using inequality (1.8), we get

$$|g_k^T d_k| \le \|g_k\|^2 + 2\mu\sigma\|g_k\|^2$$

which means

$$d_k \ge -C_1\|g_k\|^2$$

where $C_1 = 1 + 2\mu\sigma$ and this completes the proof.

Theorem 2.3: *Suppose that* Assumption 2.1 holds. *Any CG method with a coefficient* β_k *satisfying (2.1) is globally convergent when it is applied under the strong Wolfe line search with* $0 < \sigma < \frac{1}{4\mu}$, *that is,*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{2.9}$$

Proof: The proof is by contradiction. It assumes that the opposite of (2.9) holds, that is, there exists a constant $\varepsilon > 0$ and an integer k_1 such that

$$\|g_k\| \ge \varepsilon, \text{for all, } \mathrm{k} \ge \mathrm{k}_1 \tag{2.10}$$

which leads to

$$\frac{1}{\|g_k\|^2} \le \frac{1}{\varepsilon^2}, \quad \text{for all } k \ge k_1 \tag{2.11}$$

From (1.5), by squaring both sides of $d_k + g_k = \beta_k d_{k-1}$, we get

$$\|d_k\|^2 = -\|g_k\|^2 - 2g_k^T d_k + (\beta_k)^2\|d_{k-1}\|^2 \tag{2.12}$$

Using (2.8), we obtain

$$\|d_k\|^2 \le -\|g_k\|^2 + 2C_1\|g_k\|^2 + (\beta_k)^2\|d_{k-1}\|^2,$$

which means

$\|d_k\|^2 \le C_3\|g_k\|^2 + (\beta_k)^2\|d_{k-1}\|^2$, where $C_3 = 2C_1 - 1$.

Substituting (2.1) and dividing both sides by $\|g_k\|^4$, we get

$$\frac{\|d_k\|^2}{\|g_k\|^4} \le \frac{C_3}{\|g_k\|^2} + \frac{\mu^2}{\|d_{k-1}\|^2}.$$

Since $\frac{1}{\|d_{k-1}\|^2} < \frac{4}{\|g_{k-1}\|^2}$ (see (1.8)), then

$$\frac{\|d_k\|^2}{\|g_k\|^4} < \frac{C_3}{\|g_k\|^2} + \frac{4\mu^2}{\|g_{k-1}\|^2}. \tag{2.13}$$

Combining (2.11) and (2.13) together, we come to

$$\frac{\|d_k\|^2}{\|g_k\|^4} < \frac{C_3 + 4\mu^2}{\varepsilon^2}, \text{for all} k \ge k_1 + 1.$$

This means

$$\frac{\|g_k\|^4}{\|d_k\|^2} > \gamma, \text{where } \gamma = \frac{\varepsilon^2}{C_3 + 4\mu^2}.$$

Then

$$\sum_{k=k_1+1}^{n} \frac{\|g_k\|^4}{\|d_k\|^2} > (n - k_1)\gamma.$$

Since

Table 1
A comparison between FR, HS, PRP, OHS, OPRP, and RMIL +.

			NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG
1	GENERALIZED WHITE & HOLST (Andrei, 2008)	2	49/233/127 74/307/177	15/113/47 16/111/53	15/98/47 Fail	14/89/49 18/112/62	14/89/45 18/115/64	20/113/59 24/141/80
2	THREE-HUMP (Molga and Smutnicki, 2005)	2	14/397/85 Fail	26/750/93 26/770/122	11/293/48 14/347/143	26/750/93 19/530/138	11/293/48 15/385/86	14/387/99 20/547/175
3	SIX-HUMP (Molga and Smutnicki, 2005)	2	13/44/30 10/54/28	5/19/13 8/47/23	5/19/13 8/47/25	5/19/13 7/40/22	5/19/13 8/45/24	5/19/13 8/50/26
4	TRECANNI (Zhu, 2004)	2	10/40/25 9/32/20	5/26/16 4/17/10	5/26/16 Fail	5/26/16 5/20/12	5/26/16 5/19/11	5/26/16 Fail
13	EXTENDED WOOD (Andrei, 2008)	4	6897/33847/16150 1115/5012/2624	123/481/282 167/736/401	106/428/243 157/697/383	51/225/128 74/397/200	57/250/142 85/458/233	118/406/254 245/982/581
14	FREUDENSTEIN & ROTH (Andrei, 2008)	4	24/88/53 194/85/47	Fail 7/41/19	7/39/19 10/52/25	7/35/18 7/45/20	7/36/19 9/52/25	8/40/20 9/52/25
15	GENERALIZED TRIDIAGONAL 2 (Andrei, 2008)	4	5/16/13 Fail	4/13/11 10/51/34	4/13/11 10/49/33	4/13/11 12/61/43	4/13/11 12/61/43	4/13/11 12/61/43
16	QP1 (Andrei, 2008)	4	205/69/45 17/74/43	6/24/14 10/54/27	7/27/16 8/43/22	7/27/16 10/52/29	7/27/16 10/52/29	7/28/17 11/55/31
17	FLETCHER (Andrei, 2008)	10	1203/5826/2809 2443/1202/587	56/256/134 73/344/173	56/256/134 73/348/174	56/256/134 66/380/175	56/256/134 51/300/137	74/307/171 105/502/253
18	GENERALIZED TRIDIAGONAL 1 (Andrei, 2008)	10	27/88/58 43/164/103	23/76/50 27/112/68	23/76/50 27/112/68	23/76/50 27/112/68	23/76/50 27/112/68	22/73/48 27/109/66
19	HAGER (Andrei, 2008)	10	11/34/31 97/314/215	12/37/32 17/60/51	12/37/32 17/60/51	12/37/32 17/60/51	12/37/32 17/60/51	12/37/32 18/65/57
20	ARWHEAD (Andrei, 2008)	10	7/27/17 13/70/36	5/22/14 8/52/24	5/22/14 9/55/26	5/22/14 9/58/28	5/22/14 8/55/26	5/22/14 9/58/28
21	GENERALIZED QUARTIC (Andrei, 2008)	10	11/222/58 47/1065/868	8/89/57 17/335/102	8/93/59 16/320/131	7/69/39 14/223/115	8/93/59 15/236/147	6/48/17 12/154/76
22	POWER (Andrei, 2008)	10	10/30/20 103/30/20	10/30/20 10/30/20	10/30/20 10/30/20	10/30/20 10/30/20	10/30/20 10/30/20	104/312/208 122/366/244
23	GENERALIZED ROSENBROCK (Andrei, 2008)	10	Fail Fail	437/1702/1000 361/1540/868	480/1847/1103 449/1840/1054	642/2271/1431 962/3409/2124	1805/5798/3804 1211/4135/2615	1173/3915/2505 1691/5709/3598
24	RAYDAN 1 (Andrei, 2008)	10	19/90/76 2806/9964/6278	17/80/67 Fail	17/80/67 Fail	17/80/67 36/199/166	17/80/67 36/196/168	20/96/81 37/200/171
		10^2	949/484/197 880/3806/1910	74/287/152 170/723/373	75/314/157 169/694/371	74/287/152 130/559/294	75/314/157 130/565/293	86/266/179 153/587/353
25	EXTENDED DENSHENB (Andrei, 2008)	10	9/31/22 17/65/42	5/19/14 8/34/21	5/19/14 8/34/21	5/19/14 9/37/23	5/19/14 9/37/23	5/19/14 10/45/29
		10^2	18/68/44 9/44/23	8/34/21 9/51/27	8/34/21 10/48/26	9/37/23 9/44/23	9/37/23 9/44/23	10/45/29 10/47/25
26	EXTENDED PENALTY (Andrei, 2008)	10	12/52/31 17/64/38	45/162/106 10/49/29	29/106/67 11/52/33	16/63/39 9/40/23	15/60/37 9/40/23	14/57/36 8/37/22
		10^2	238/4732/506 2830/11510/4309	17/97/51 15/96/49	Fail Fail	13/78/40 12/71/37	13/81/42 12/71/37	23/144/78 13/75/39
27	QP2 (Andrei, 2008)	10^2	283/3487/737 2482/3295/649	21/219/74 27/285/88	23/244/77 24/256/84	35/322/114 31/294/100	36/325/116 32/308/104	33/301/105 30/287/97
28	DIXON3DQ (Andrei, 2008)	50	25/79/55 25/79/55	25/79/55 25/79/55	25/79/55 25/79/55	25/79/55 25/79/55	25/79/55 25/79/55	123/376/262 127/392/275
29	QF2 (Andrei, 2008)	50	116/394/285 73/299/168	70/244/160 66/272/150	70/244/160 65/269/148	70/244/160 66/272/150	70/244/160 65/269/148	78/274/181 74/305/177
30	QF1 (Andrei, 2008)	50	38/114/76 40/120/80	38/114/76 40/120/80	38/114/76 40/120/80	38/114/76 40/120/80	38/114/76 40/120/80	69/207/138 78/234/156
		500	131/393/262 137/411/274	131/393/262 137/411/274	131/393/262 137/411/274	131/393/262 137/411/274	13/393/262 137/411/274	162/486/325 198/594/397
31	HIMMELH (Andrei, 2008)	500	13/79/29 11/64/23	5/15/10 5/15/10	5/15/10 5/15/10	6/32/13 6/18/12	6/32/13 6/18/12	5/15/10 6/18/12
32	QUARTC (Andrei, 2008)	500	3/31/26 4/43/33	Fail Fail	2/24/23 Fail	3/31/26 3/26/16	3/31/26 3/26/16	3/31/26 3/26/16
33	EXTENDED TRIDIAGONAL 1 (Andrei, 2008)	500	340/1190/1035 14/68/58	14/71/60 7/42/35	14/72/58 11/63/53	12/61/50 8/47/38	12/61/50 8/47/38	12/60/50 8/47/38
		10^3	399/1396/1212 518/1813/1561	14/71/60 7/43/25	14/72/58 14/83/61	12/61/51 13/78/55	12/61/50 13/77/55	12/60/50 7/44/24
34	DIAGONAL 4 (Andrei, 2008)	500	2/6/5 2/6/5	2/6/5 2/6/5	2/6/5 2/6/5	2/6/5 2/6/5	2/6/5 2/6/5	2/6/5 2/6/5
		10^3	2/6/5 2/6/5	2/6/5 2/6/5	2/6/5 2/6/5	2/6/5 2/6/5	2/6/5 2/6/5	2/6/5 2/6/5
35	EXTENDED WHITE & HOLST (Andrei, 2008)	500	57/257/143 282/3749/859	15/113/47 50/585/212	15/98/47 49/548/207	15/95/50 50/432/195	15/95/50 50/432/194	22/121/65 52/436/199
		10^3	572/257/143 1231/1608/372	15/113/47 35/297/135	15/98/47 35/292/128	15/95/50 48/371/168	15/95/50 49/376/170	22/121/65 120/914/418
36	EXTENDED ROSENBROCK (Andrei, 2008)	10^3	68/530/182 260/2899/718	19/120/58 23/176/70	21/134/67 25/183/72	28/168/90 33/219/102	28/167/88 32/215/99	31/181/99 27/185/88
		10^4	71/539/188	19/120/58	21/134/67	28/168/90	28/167/88	31/181/99

Table 1 (*continued*)

			NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG
			715/6694/1764	17/115/53	18/117/54	18/105/57	18/105/57	45/301/148
37	EXTENDED HIMMELBLAU (Andrei, 2008)	10^3	15/54/34	7/29/17	8/32/19	8/32/19	8/32/19	7/30/18
			11/51/27	9/48/24	9/43/22	7/39/19	7/39/19	7/39/19
		10^4	22/127/55	9/44/23	9/44/23	10/47/25	9/44/23	9/39/22
			17/72/38	10/52/24	Fail	9/47/21	9/47/21	9/48/22
38	STRAIT (Mishra, 2005)	10^3	35/146/91	17/86/48	17/86/48	15/80/44	15/80/44	20/96/55
			88/682/269	19/148/59	19/150/60	19/179/71	18/172/63	22/193/78
		10^4	35/147/92	18/90/51	18/90/51	15/80/44	15/80/44	20/97/55
			109/774/348	20/129/59	20/127/60	19/185/71	18/184/68	23/167/71
39	SHALLOW (Issam, 2005)	10^3	18/63/48	7/27/21	7/17/21	7/28/22	7/28/22	7/28/22
			179/605/390	13/63/42	14/66/39	12/62/39	12/62/39	15/70/46
		10^4	46/144/97	8/28/20	8/28/20	9/35/26	9/35/26	8/33/25
			19/71/49	9/38/26	10/41/28	10/49/32	10/49/32	10/47/31
40	EXTENDED BEALE (Andrei, 2008)	10^3	75/242/159	10/48/31	13/67/41	11/52/33	14/72/46	14/62/43
			80/254/170	10/42/29	11/45/31	12/52/39	12/52/39	13/53/40
		10^4	88/285/187	9/41/26	9/41/24	10/48/31	9/43/26	6/30/18
			86/272/182	10/42/29	11/45/31	12/52/39	12/52/39	13/53/40

$$\sum_{k=0}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2} > \sum_{k=k_1+1}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2} = \lim_{n\to\infty}\sum_{k=k_1+1}^{n}\frac{\|g_k\|^4}{\|d_k\|^2},$$

and

$$\sum_{k=k_1+1}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2} = \lim_{n\to\infty}\sum_{k=k_1+1}^{n}\frac{\|g_k\|^4}{\|d_k\|^2} > \lim_{n\to\infty}(n-k_1)\gamma = \infty.$$

We come to

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

This contradicts (2.7). Therefore, the proof is completed.

3. Modified versions of the PRP and the HS methods

In this section, since the sufficient descent property and the global convergence of the well-known PRP and HS methods are not established under strong Wolfe line search, then motivated by the results in Section 2, we propose modified versions of PRP and HS methods, that is, by restricting the coefficients β_k^{PRP} and β_k^{HS} in order to satisfy (2.1) as follows

$$\beta_k^{OPRP} = \begin{cases} \beta_k^{PRP} & \text{if} \quad -\mu\frac{g_k^2}{d_{k-1}^2} < \beta_k^{PRP} < \mu\frac{g_k^2}{d_{k-1}^2} \\ 0, & \text{otherwise}. \end{cases} \tag{3.1}$$

and

$$\beta_k^{OHS} = \begin{cases} \beta_k^{HS} & \text{if} \quad -\mu\frac{g_k^2}{d_{k-1}^2} < \beta_k^{HS} < \mu\frac{g_k^2}{d_{k-1}^2} \\ 0, & \text{otherwise}. \end{cases} \tag{3.2}$$

We call these modified versions OPRP and OHS respectively, where the letter "O" stands for Osman.

Of course, both of the modified versions of PRP and HS satisfy (2.1), so that they generate descent directions at each iteration and globally convergent when they are applied under strong Wolfe line search with $0 < \sigma < \frac{1}{4\mu}$. Note that, in (3.1) and (3.2) when $\mu \to \infty$, then $\beta_k^{OPRP} \to \beta_k^{PRP}$ and $\beta_k^{OHS} \to \beta_k^{HS}$ and also σ tends to zero. Therefore, for a sufficiently large value of μ, the OPRP and the OHS methods can be considered as good approximations to both PRP and HS methods.

We also note, like PRP and HS methods, OPRP and OHS methods perform a restart when they encounter a bad direction, i.e., when g_k approaches g_{k-1}, then both β_k^{OPRP} and β_k^{OHS} approach zero, so that d_k approaches $-g_k$. Hence, we expect that they perform better than FR method in practice. Also, the sufficient descent property and the global convergence of both OPRP and OHS methods qualified them to be better than both PRP and HS theoretically, but it remains to show their performance in practical computations and this will be shown in the next section.

4. Numerical experiment

In this section, to show the efficiency and robustness and to support the theoretical proofs in Section 2, numerical experiments based on comparing the proposed OPRP and OHS when $\mu = 10$ with PRP, HS, FR, and RMIL + methods are done. To accomplish the comparison, a MATLAB coded program for these methods when they are all implemented under strong Wolfe line search with $\delta = 10^{-4}$ and $\sigma = 10^{-2}$ is run. We stop the program if $\|g_k\| \le 10^{-6}$. The test problems are unconstrained and most of them are from (Andrei, 2008). To show the robustness, test problems are implemented under low, medium, and high dimensions, namely, 2, 3, 4, 10, 50, 100, 500, 1000, and 10000. Furthermore, for each dimension, two different initial points are used, one of them is the initial point, which is suggested by Andrei (Andrei, 2008) and the other point is chosen arbitrarily. The comparison is based on the number of iterations (NI), the number of function evaluations (NF), and the number of gradient evaluations (NG). The numerical results are in Table 1. In Table 1, a method is considered to have failed, and we report "Fail" if the number of iterations exceeds 5×10^3, or the search direction is not descent.

According to Table 1, we show the performance of OPRP, OHS, HS, PRP, FR, and RMIL + methods in Figs. 1-3 relative to the number of iterations, number of function evaluations, and number of gradient evaluations respectively. We used the performance profile introduced by Dolan and More (Dolan and More, 2002) which provides solver efficiency, robustness, and probability of success. In Dolan and More performance profile, we plot the percentage P of the test questions where a method falls within the best *t*-factor. Obviously, in the performance profile table, the curved shape at the top of the method is the winner. Furthermore, plot correctness is a measure of the robustness of the method.

Clearly, from Figs. 1-3, OPRP and OHS solve all test problems and therefore, reach 100 % percentage, whereas, FR, HS, PRP, and RMIL + solve about 94 %, 96 %, 90 %, and 99 % respectively. Furthermore, the left sides of all figures show that OPRP, OHS, PRP, and HS almost have the same highest probability of being the optimal sol-

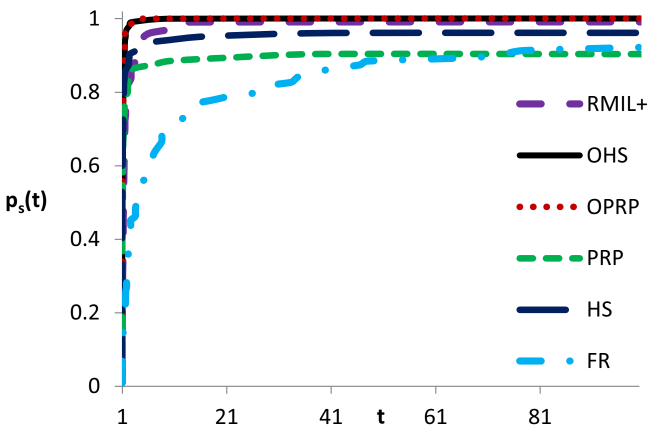

Fig. 1. The performance based on NI.

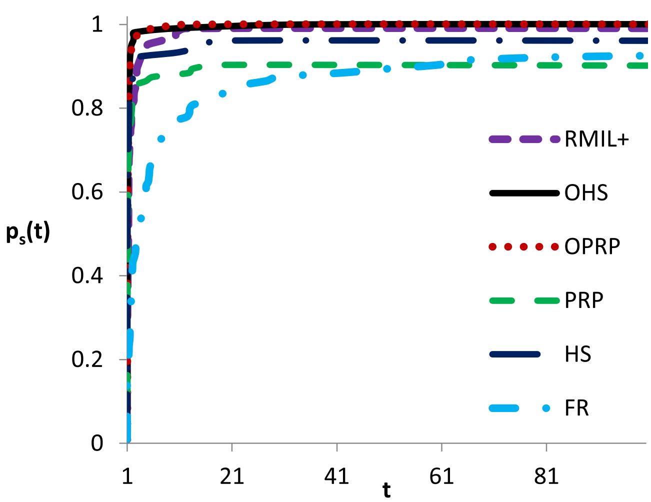

Fig. 2. The performance based on NF.

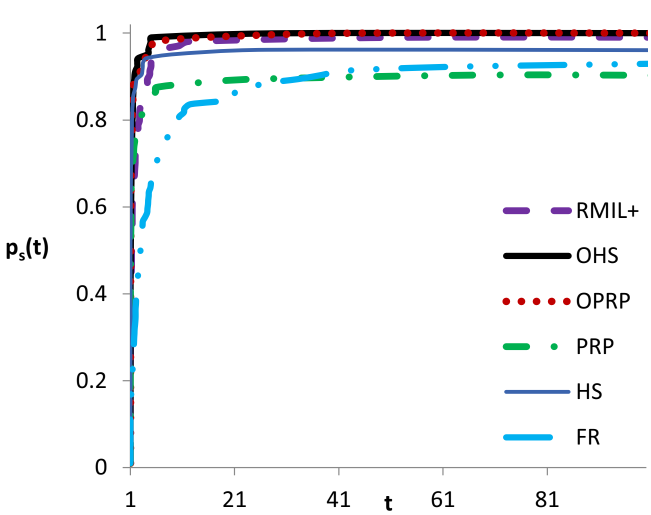

Fig. 3. The performance based on NG.

vers. In general, since the curves of both OPRP and OHS are above all other curves in most cases, then their performance is better than of the other methods.

5. Conclusions

In this paper, under the strong Wolfe line search, with $0 < \sigma < \frac{1}{4\mu}$, $\mu \geq 1$, we established the sufficient descent property and global convergence of CG methods with their coefficient β_k satisfying $|\beta_k| \leq \mu \frac{\|g_k\|^2}{\|d_{k-1}\|^2}$, for all $k \geq 1$. At the same time, we have proposed new modified versions of both PRP and HS methods called OPRP and OHS respectively. To show the efficiency and robustness and to support the theoretical proofs which establish the sufficient descent property and the global convergence, numerical experiments based on comparing OPRP and OHS with HS, PRP, FR, and RMIL+ have been done. Based on Dolan and More performance profile, it has been found that the new modified versions perform better than the other methods.

Funding

This research is supported by the deanship of scientific research in Majmaah University under the project number R-2022-230.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Mogtaba Mohammed - Majamaah University, Osman Yousif - University of Gezera, Mohammed Saleh - Qassim University, Murtada Elbashir - Jouf University.

References

Abubakar, A.B., Malik, M., Kumam, P., Mohammad, H., Sun, M., Ibrahim, A.H., Kiri, A. I., 2022. A Liu-Storey-type conjugate gradient method for unconstrained minimization problem with application in motion control. J. King Saud Univ. Sci. 34, (4) 101923.
Al-Baali, M., 1985. Descent property and global convergence of the Fletcher-Reeves method with inexact line search. IMA J. Numer. Anal. 5, 121–124.
Andrei, N., 2008. An unconstrained optimization test functions collection. Adv. Model. Pptimizat. 10 (1), 147–161.
Dai, Y., 2002. A nonmonotone conjugate gradient algorithm for unconstrained optimization. J. Syst. Sci. Complex. 15 (2), 139–145.
Dai, Z., 2016. Comments on a new class of nonlinear conjugate gradient coefficients with global convergence properties. Appl. Math. Comput 276, 297–300.
Dai, Y.H., Yuan, Y., 2000. A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optim. 10, 177–182.
Dolan, E., More, J.J., 2002. Benchmarking optimization software with performance profile. Math. Prog. 91, 201–213.
Fletcher, R., 1987. Practical Method of Optimization. Unconstrained Optimization Vol. 1.
Fletcher, R., Reeves, C., 1964. Function minimization by conjugate gradients. Comput. J. 7, 149–154.
Gilbert, J.C., Nocedal, J., 1992. Global convergence properties of conjugate gradient methods for optimization. SIAM. J. Optim. 2, 21–42.
Hager, W.W., Zhang, H., 2005. A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM. J. Optim. 16 (1), 170–192.
Hestenes, M.R., Stiefel, E.L., 1952. Methods of conjugate gradients for solving linear systems. J. Research Nat. Bur. Standards. 49, 409–436.
Hu, Y.F., Storey, C., 1991. Global convergence result for conjugate gradient methods. J. Optim. Theory Appl. 71, 399–405.
Issam, A.R.M., 2005. Minimization of extended quadratic functions with inexact line searches. J. Korean Soc. Indust. Appl. Math. 9 (1), 55–61.
Liu, Y., Storey, C., 1992. Efficient generalized conjugate gradient algorithms. Part 1: Theory. J. Optimiz. Theory Appl. 69, 129–137.
Mishra, S., 2005. Some New Test Functions for Global Optimization and Performance of Repulsive Particle Swam Method. North-Eastern Hill University, India.
M. Molga and C. Smutnicki, Test Functions for Optimization Needs, www.zsd.ict.pwr.wroc.p1/files/docs/functions.pdf, 2005.
Polak, E., Ribière, G., 1969. Note sur la convergence de directions conjuguée. Rev. Francaise Informat Recherche Operationelle, 3e Année 16, 35–43.
Polyak, B.T., 1969. The conjugate gradient method in extremem problems. USSR Comp. Math. Math. Phys. 9, 94–112.
Powell, M.J.D., 1984. Nonconvex Minimization Calculations And The Conjugate Gradient Method In Lecture Notes In Mathematics, 1066. Springer-Verlag, Berlin, pp. 122–141.

Powell, M.J.D., 1986. Convergence properties of algorithms for nonlinear optimization. SIAM Rev. 28, 487–500.
Rivaie, M., Mamat, M., Leong, W.J., Mod, I., 2012. A new class of nonlinear conjugate gradient coefficient with global convergence properties. Appl. Math. Comput. 218, 11323–11332.
Salleh, Z., Almarashi, A., Alhawarat, A., 2022. Two efficient modifications of AZPRP conjugate gradient method with sufficient descent property. J. Inequal. Appl. 1, 1–21.
Touati-Ahmed, D., Storey, C., 1990. Efficient hybrid conjugate gradient techniques. J. Optim. Theory Appl. 64, 379–397.
Wei, Z., Li, G., Qi, L., 2006. New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems. Appl. Math. Comput 179 (2), 407–430.
Wei, Z., Yao, S., Liu, L., 2006. The convergence properties of some new conjugate gradient methods. Appl. Math. Compu. 183 (2), 1341–1350.
Yousif, O.O.O., 2020. The convergence properties of RMIL+ conjugate gradient method under strong Wolfe line search. Appl. Math. Compu. 367, 124777.
Yuan, G., Lu, X., 2009. A modified PRP conjugate gradient method. Ann. Oper. Res. 166, 73–90.
Yuan, Y., Sun, W., 1999. Theory and Methods of Optimization. Science Press of China, Beijing, China.
Zhang, L., 2009. An improved Wei-Yao-Liu nonlinear conjugate gradient method for optimization computation. Appl. Math. Comput. 215, 2269–2274.
Zhu, W.X., 2004. A Class of Filled Functions irrelevant to the Number of Local Minimizers for Global Optimization. J. Syst. Sci. Math. Sci. 22 (4), 406–413.
Zoutendijk, G., 1970. Nonlinear programming, computational methods. In: Abadie, J. (Ed.), Integer and Nonlinear Programming. North- Holland, Amsterdam, pp. 37–86.