TWMS J. App. and Eng. Math. V.9, N.3, 2019, pp. 535-548

A DESCENT PRP CONJUGATE GRADIENT METHOD FOR UNCONSTRAINED OPTIMIZATION

H. NOSRATIPOUR[1], K. AMINI[1], §

Abstract. It is well known that the sufficient descent condition is very important to the global convergence of the nonlinear conjugate gradient methods. Also, the direction generated by a conjugate gradient method may not be a descent direction. In this paper, we propose a new Armijo-type line search algorithm such that the direction generated by the PRP conjugate gradient method has the sufficient descent property and ensures the global convergence of the PRP conjugate gradient method for the unconstrained minimization of nonconvex differentiable functions. We also present some numerical results to show the efficiency of the proposed method.The results show the efficiency of the proposed method in the sense of the performance profile introduced by Dolan and Moré.

Keywords: Unconstrained optimization, Armijo-type line search, Conjugate gradient method, sufficient descent, Global convergence.

AMS Subject Classification: 90C30, 65K05

1. Introduction

The nonlinear conjugate gradient (CG) method plays a very important role for solving the unconstrained optimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. CG methods are usually designed by the iterative form

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

where x_k is the current iterate point, $\alpha_k > 0$ is a steplength and d_k is the search direction defined by

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

[1] Razi University, Faculty of Science, Department of Mathematics, Kermanshah, Iran.
e-mail: hadi.nosratipour@gmail.com; ORCID: https://orcid.org/0000-0001-8801-9301.
e-mail: kamini@razi.ac.ir; ORCID: https://orcid.org/0000-0002-9264-3091.

§ Manuscript received: May 13, 2017; accepted: October 10, 2017.
TWMS Journal of Applied and Engineering Mathematics, Vol.9, No.3 © Işık University, Department of Mathematics, 2019; all rights reserved.

where $\beta_k \in \mathbb{R}$ is known as the conjugate gradient parameter. A lot of versions of conjugate gradient methods, correspond to the selection procedure of parameters β_k, are already known. Some of these selections are given as follows.

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \tag{4}$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{(g_k - g_{k-1})^T d_{k-1}}, \tag{5}$$

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}, \tag{6}$$

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}, \tag{7}$$

where $\|.\|$ denotes the Euclidian norm [8, 12, 11, 17, 2, 5]. Although these methods are equivalent when f is a strictly convex quadratic function and α_k is computed by the following exact line search rule, i.e.

$$f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k), \tag{8}$$

but their behaviors for general objective functions may be far different. Their convergence properties have been studied by many authors, including [8, 12, 11, 17, 2, 5].
The global convergence of the PRP method is established when f is strongly convex and the line search is exact [13]. Powell [15] has proved that for a general nonlinear function, the PRP is globally convergent if

(1) the stepsize $x_{k+1} - x_k$ approaches zero,
(2) the line search is exact,
(3) the Lipschitz Assumption of g holds.

In 1984, by using of a 3 dimensional example, Powell [14] showed that even with α_k obtained by an exact line search, the PRP method can cycle infinitely without approaching to a stationary point. Hence, this assumption that the stepsize tends to zero is needed for convergence. Under the assumption that the search direction is a descent direction, Yuan [20] has established the global convergence of the PRP method for strongly convex objective functions along with the Wolfe line search. However, for the strong Wolfe line search, Dai [3] has introduced an example which shows that even when the objective function is strongly convex and $\sigma \in (0, 1)$ (the parameter of curvature condition) is sufficiently small, the PRP method may still fail by generating an ascent search direction.

In summary, the convergence of the PRP method for general nonlinear function is uncertain, Powell's example shows that when the function is not strongly convex, the PRP method may not converge even with an exact line search. Based on insight gained from this example and to cope with possible convergence failure in the PRP algorithm, Powell [15] has suggested the following modification in the update parameter for the PRP method as

$$\beta_k^{PRP^+} = \max\{\beta_k^{PRP}, 0\}. \tag{9}$$

Gilbert and Nocedal [9] have shown that this choice guarantees the global convergence of the PRP$^+$ algorithm. However, for the unconstrained minimization of strongly convex (nonquadratic) functions with exact line searches, Gilbert and Nocedal showed that β_k^{PRP} can be negative when the PRP method is convergent. This indicates that the choice $\beta_k^{PRP^+}$ may not be the only way to enforce global convergence of the PRP method. Although the PRP$^+$ algorithm addresses the possibility of convergence failure, it interferes with the n-step convergence property of the conjugate gradient method for strongly convex quadratic

functions and possibly decreases the rapid convergence of algorithm.
Another important approach for rectifying the convergence failure in the PRP algorithm is to retain the PRP update formula and modify the line search. In this regard, Grippo and Lucidi [10] by modifying Leone et al.'s line search [6] proposed an new line search as follows.

Given constants $\tau > 0$, $\sigma \in (0,1)$, $\delta > 0$ and $0 < c_1 < 1 < c_2$, their line search aims to find

$$\alpha_k = \max\{\sigma^j \tau \frac{|g_k^T d_k|}{\|d_k\|^2};\ j = 0, 1, \cdots\}, \tag{10}$$

such that $x_{k+1} = x_k + \alpha_k d_k$ with $d_{k+1} = -g_{k+1} + \beta_{k+1}^{PRP} d_k$ satisfies

$$f(x_{k+1}) \leq f(x_k) - \delta \alpha_k^2 \|d_k\|^2, \tag{11}$$

and

$$-c_2\|g_{k+1}\|^2 \leq g_{k+1}^T d_{k+1} \leq -c_1\|g_{k+1}\|^2. \tag{12}$$

Grippo and Lucidi [10] have proved the global convergence of the PRP conjugate gradient method equipped the line search rule (10)-(12) for the unconstrained minimization of nonconvex differentiable functions.
Dai [4] also has proposed a new line search for nonlinear conjugate gradient methods as follows.

Given $\lambda \in (0,1)$, $\delta \in (0,1)$ and $c_1 \in (0,1)$, determine the smallest integer $m \geq 0$ such that if one defines

$$\alpha_k = \lambda^m, \tag{13}$$

then

$$f(x_k + \alpha_k d_k) \leq f_k + \delta \alpha_k g_k^T d_k, \tag{14}$$

$$0 \neq g_{k+1}^T d_{k+1} \leq -c_1\|d_{k+1}\|^2. \tag{15}$$

The global convergence of the PRP conjugate gradient method with this line search rule has been proved in [4].

The PRP method generally performs better than the other conjugate gradient methods in practice. However, it is not generally a descent method when Armijo-type line search is used, thus [10] and [4] for satisfying sufficient descent property added the extra relation (12) and (15) to Armijo-type line search, respectively. These strategies are valuable from the theoretical viewpoint.

It has been proved that these line search approaches are well defined and have the advantage that they can guarantee the global convergence of the original PRP method. However similarly to the strong Wolfe line search, they are computationally expensive. More precisely, to calculate a steplength to satisfy in the second condition any of these strategies (i.e., the inequalities (12) and (15)), it may be necessary to compute g_{k+1} and d_{k+1} several times in each iteration. Hence, it is clear that the above mentioned line search rules are more computationally expensive.

The purpose of this paper is to overcome this drawback. To this end, using an estimated local Lipschitz constant of the derivative of objective function and choosing an adequate initial steplength s_k, we modify and improve the two line search strategies proposed in [4, 10] for computing a suitable steplength α_k at each iteration by omitting the second inequalities (12) and (15). We prove that, under some mild conditions, the PRP method along with the new line search rules generate search directions that satisfy the sufficient descent condition. Also, the PRP method is proved to be strong globally convergent.

The paper is organized as follows. In §2, we present a new Armijo-type line search. In §3, we prove the global convergence of the proposed algorithm. In §4, to show efficiency of the adaptive initial steplength s_k and the new line search rules, we reported some numerical results.

Notation. Throughout this paper $g(x) = \nabla f(x)$ denotes the gradient of $f(x)$. We write $\|.\|$ for the Euclidean norm of a vector. Furthermore, For all values, a subscript k means that this is the evaluation at x_k or the value in the kth iteration, e.g., f_k, g_k.

2. New algorithm

In this section, we propose a modified Armijo-type line search that is used in conjunction with the PRP method. Its form is similar to that is given in relation (11) by Grippo and Lucidi but with different initial steplength s_k which allows us to establish a global convergence result. Throughout this paper, we consider the following assumptions in order to analyze the new algorithm:

(H1). The function f is continuously differentiable and bounded below on the level set $L(x_0) = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$.

(H2) The gradient $g(x)$ of $f(x)$ is Lipschitz continuous over an open convex S that contains $L(x_0)$; i.e., there exists a positive constant L such that

$$\|g(x) - g(y)\| \leq L\|x - y\|,$$

for all $x, y \in S$.

Considering the points listed in beginning of this section, we now present Algorithm 1 to describe the steps of the PRP method with a new Armijo-type line search as follows.

In the following, we assume an infinite sequence $\{x_k\}$ is generated, otherwise, Algorithm 1 stops at a stationary point of problem (1).

Lemma 2.1. *Suppose that (H2) holds and Algorithm 1 generates an infinite sequence $\{x_k\}$, then there exists $c \in (0, 1)$, such that*

$$(c - 2)\|g_k\|^2 \leq g_k^T d_k \leq -c\|g_k\|^2. \tag{16}$$

Proof. By induction, for $k = 0$, $g_0^T d_0 = -\|g_0\|^2$, since $0 < c < 1$, we have

$$(c - 2)\|g_0\|^2 \leq -\|g_0\|^2 \leq c\|g_0\|^2,$$

so (16) holds. For $k \geq 1$, from (3) and (6) we have

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

Using of hypothesis induction, we have $g_{k-1}^T d_{k-1} < 0$, then

$$\begin{aligned} |g_k^T d_k + \|g_k\|^2| &= |\beta_k^{PRP} g_k^T d_{k-1}| \\ &\leq \frac{\|g_k\|^2 \|g_k - g_{k-1}\|}{\|g_{k-1}\|^2} \|d_{k-1}\| \\ &\leq \frac{L\alpha_{k-1}\|d_{k-1}\|^2}{\|g_{k-1}\|^2} \|g_k\|^2 \\ &\leq \frac{Ls_{k-1}\|d_{k-1}\|^2}{\|g_{k-1}\|^2} \|g_k\|^2. \end{aligned}$$

Algorithm 1: PRP method with an new Armijo-type line search

Input: $x_0 \in \mathbb{R}^n$, constants $\delta \in (0,1)$, $\rho \in (0,1)$, $c \in (0,1)$, and a stoping tolerance $\epsilon > 0$;
Output: x_b, f_b;
begin
  compute f_0 and g_0;
  $k \leftarrow 0$;
  while $\|g_k\| \geq \epsilon$ **do**
    compute d_k by (3) and (6);
    set $s_k \leftarrow \frac{1-c}{L} \frac{\|g_k\|^2}{\|d_k\|^2}$;
    $\alpha \leftarrow s_k$;
    $\widehat{x}_k \leftarrow x_k + \alpha d_k$;
    while $f(\widehat{x}_k) > f_k - \alpha^2 |d_k\|^2$ **do**
      $\alpha \leftarrow \rho\alpha$;
      $\widehat{x}_k \leftarrow x_k + \alpha d_k$;
    end
    $x_{k+1} \leftarrow \widehat{x}_k$; $f_{k+1} \leftarrow f(\widehat{x}_k)$;
    compute g_{k+1};
    $k \leftarrow k+1$;
  end
  $x_b \leftarrow x_k$; $f_b \leftarrow f_k$;
end

So, by the definition of s_k, we have

$$|g_k^T d_k + \|g_k\|^2| \leq (1-c)\|g_k\|^2$$

and

$$(c-2)\|g_k\|^2 \leq g_k^T d_k \leq -c\|g_k\|^2.$$

Thus (16) holds and the proof is complete. □

Lemma 2.2. *Suppose that (H2) holds and Algorithm 1 generates an infinite sequence $\{x_k\}$, then there exists $c \in (0,1)$, such that*

$$\|d_k\| \leq (2-c)\|g_k\|. \tag{17}$$

Proof. For $k = 0$, $\|d_0\| = \|g_0\|$, and so (17) holds. For $k \geq 1$, from (3) and (6) we have

$$\begin{aligned}\|d_k\| &= \|-g_k + \beta_k^{PRP} d_{k-1}\| \\ &\leq \|g_k\| + \frac{\|g_k\|\|g_k - g_{k-1}\|}{\|g_{k-1}\|^2}\|d_{k-1}\| \\ &\leq (1 + \frac{L\alpha_{k-1}\|d_{k-1}\|^2}{\|g_{k-1}\|^2})\|g_k\| \\ &\leq (1 + \frac{Ls_{k-1}\|d_{k-1}\|^2}{\|g_{k-1}\|^2})\|g_k\|.\end{aligned}$$

So, by the definition of s_k, we have

$$\|d_k\| \leq (2-c)\|g_k\|.$$

Thus (17) holds and the proof is complete. □

Lemma 2.3. *Suppose that (H1) and (H2) hold, then the line search (11) is well-defined.*

Proof. By Taylor theorem, we can observe

$$f(x_k + \alpha d_k) = f_k + \alpha g_k^T d_k + O(\alpha^2 \|d_k\|^2).$$

So, we deduce

$$\begin{aligned}&\lim_{\alpha \to 0^+} \frac{f_k - f(x_k + \alpha d_k) - \delta \alpha^2 \|d_k\|^2}{\alpha} \\ &= \lim_{\alpha \to 0^+} \frac{-\alpha g_k^T d_k - O(\alpha^2 \|d_k\|^2) - \delta \alpha^2 \|d_k\|^2}{\alpha} \\ &= -g_k^T d_k > 0.\end{aligned}$$

Since $\alpha > 0$, it is concluded that there exists $\acute{\alpha}_k > 0$ such that

$$f(x_k + \alpha d_k) \le f_k - \delta \alpha^2 \|d_k\|^2, \quad \forall \alpha \in [0, \acute{\alpha}_k].$$

Setting $\hat{\alpha}_k = \min(s_k, \acute{\alpha}_k)$ yields

$$f(x_k + \alpha d_k) \le f_k - \delta \alpha^2 \|d_k\|^2, \quad \forall \alpha \in [0, \hat{\alpha}_k].$$

So, the new line search is well-defined and the proof is complete. □

3. Convergence analysis

In this section, we prove the global convergence of Algorithm 1 under mild assumptions.

Lemma 3.1. *Suppose that (H1) and (H2) hold and Algorithm 1 generates an infinite sequence $\{x_k\}$. then*

$$\eta_0 = \inf_{\forall k \ge 0} \{\alpha_k\}, \tag{18}$$

is positive.

Proof. On the contrary, we suppose $\eta_0 = 0$. So, there exists an infinite subset $K \subseteq \{0, 1, 2, ...\}$ such that

$$\lim_{k \in K\ ,k \to \infty} \alpha_k = 0. \tag{19}$$

By the definition of s_k and (17), we have

$$s_k \ge \frac{1-c}{L(2-c)^2} > 0. \tag{20}$$

Thus, the sequence $\{s_k\}$ is positive and bounded from below. This along with (19) imply that there is $\acute{k}$ such that

$$\alpha_k / \rho \le s_k \qquad \forall k \ge \acute{k} \ \ \textit{and} \ \ k \in K.$$

On the other hand, from the line search rule (11), we observe that $\hat{\alpha} = \alpha_k/\rho$ doesn't satisfy (11), so

$$f(x_k + \hat{\alpha} d_k) > f_k - \delta \hat{\alpha}^2 \|d_k\|^2. \tag{21}$$

This leads to

$$f(x_k + \hat{\alpha} d_k) - f_k > -\delta \hat{\alpha}^2 \|d_k\|^2. \tag{22}$$

Using the mean value theorem on the left-hand side of the above inequality, there exists $\theta_k \in (0, 1)$ such that

$$f(x_k + \hat{\alpha} d_k) - f_k = \hat{\alpha} g(x_k + \hat{\alpha} \theta_k d_k)^T d_k,$$

so

$$-\delta\hat{\alpha}^2\|d_k\|^2 < \hat{\alpha}g(x_k + \hat{\alpha}\theta_k d_k)^T d_k.$$

Thus, we can deduce

$$-\delta\hat{\alpha}\|d_k\|^2 < g_k^T d_k + (g(x_k + \hat{\alpha}\theta_k d_k) - g_k)^T d_k. \tag{23}$$

By the Cauchy-Schwartz inequality and (23), we have

$$-\delta\hat{\alpha}\|d_k\|^2 < g_k^T d_k + \|g(x_k + \hat{\alpha}\theta_k d_k) - g_k\|\|d_k\|. \tag{24}$$

Using of (H2) along with $\theta_k \in (0, 1)$, we have

$$-\delta\hat{\alpha}\|d_k\|^2 < g_k^T d_k + L\hat{\alpha}\theta_k\|d_k\|^2 < g_k^T d_k + L\hat{\alpha}\|d_k\|^2. \tag{25}$$

Thus, we obtain

$$\hat{\alpha} \geq \frac{1}{L+\delta}\frac{-g_k^T d_k}{\|d_k\|^2}. \tag{26}$$

Since d_k is a descent direction, so we have

$$\alpha_k = \hat{\alpha}\rho \geq \frac{\rho}{L+\delta}\frac{|g_k^T d_k|}{\|d_k\|^2}. \tag{27}$$

It follows from (27), Lemma 2.1 and Lemma 2.2 that

$$\alpha_k \geq \frac{\rho c}{(L+\delta)(2-c)^2} > 0 \quad for\ k \geq \acute{k}. \tag{28}$$

This contradicts (19). Therefore the proof is complete. □

Theorem 3.1. *Suppose that (H1) and (H2) hold and Algorithm 1 generates an infinite sequence $\{x_k\}$. Then*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{29}$$

Proof. Cauchy-Schwartz inequality and Lemma 2.1 along with $g_k^T d_k \leq 0$ imply that

$$\|d_k\| \geq c\|g_k\|. \tag{30}$$

By Lemma 3.1, (11) and (30), we have

$$f_k - f_{k+1} \geq \delta\alpha_k^2\|d_k\|^2 \geq \delta\eta_0^2 c^2\|g_k\|^2. \tag{31}$$

This accompanied by assumption (H1) get

$$\sum_{k\geq 0} \|g_k\|^2 < +\infty.$$

Thus (29) is hold and the proof is complete. □

4. Discussion

In Algorithm 1, we have proposed an Armijo-type line search rule equipped an adaptive initial steplength to be used in conjunction with the PRP method and established some convergence results. We note the proposed Armijo-type line search rule base on

$$f(x_{k+1}) \leq f(x_k) - \delta\alpha_k^2\|d_k\|^2, \tag{32}$$

is also suitable for free-derivative methods. In comparison with the usual acceptance condition

$$f(x_k + \alpha_k d_k) \leq f_k + \delta\alpha_k g_k^T d_k, \tag{33}$$

the condition (32) enforces a greater reduction of f, for large values of $\alpha_k\|d_k\|$, while it becomes more tolerant when $\alpha_k\|d_k\|$ is small. In practical computations, to make the steplength α_k easily accepted, it may be very useful to use the following *ad hoc* condition

$$f(x_k + \alpha_k d_k) \leq f_k + \max\{\delta\alpha_k g_k^T d_k, -\gamma\alpha_k^2\|d_k\|^2\}, \tag{34}$$

where δ and γ are constants in (0, 1)[4]. In fact, in the right hand side of (34), the max term makes steplength to be more easily accepted than the acceptance conditions (32) and (33).

we wonder whether, when the condition (32) in Algorithm 1 is replaced to the condition (33) or the condition (34), the similar convergence results can be established. The answer is Yes, and it is interesting that the theoretical results obtained in previous Lemmas and Theorem 3.1 are still true. Moreover, if initial steplength in Algorithm 1 is replaced to

$$s_k = \rho_k \frac{|g_k^T d_k|}{\|d_k\|^2}, \tag{35}$$

proposed in [10] with $\rho_k = \frac{1-c}{L}$, for all k, then the theoretical results obtained in previous Lemmas and Theorem 3.1 are still true.

5. Numerical results

In this section, we report some computational performances of the new algorithm on a set including 75 unconstrained optimization test problems. The test problems and initial points have been selected from Andrei collection of unconstrained test functions [1]. We have performed our experiments in double precision arithmetic format in MATLAB

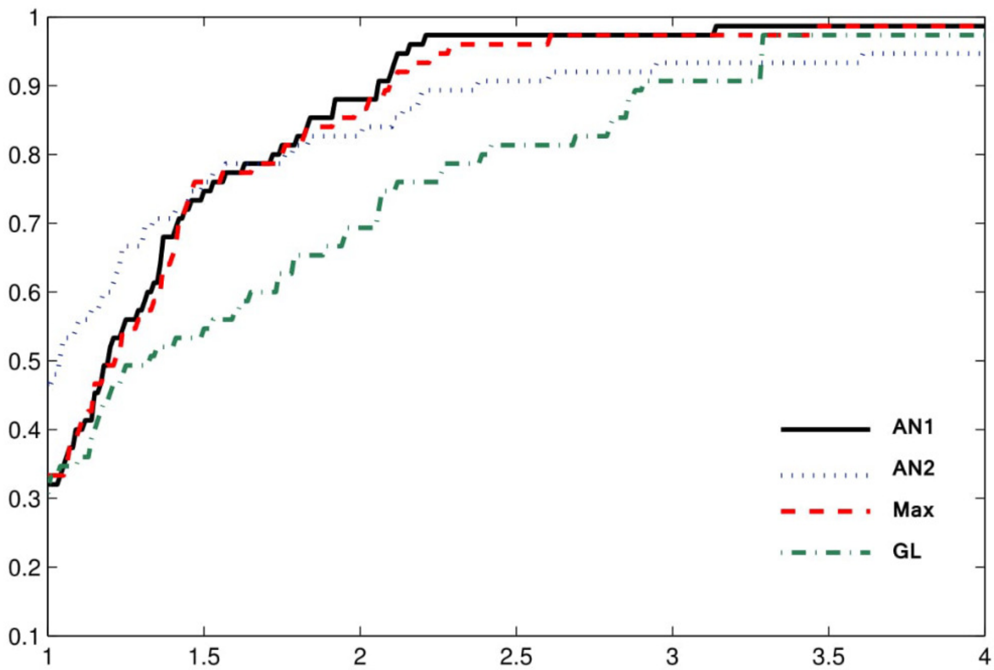

Figure 1. Iteration performance profiles for the four algorithms

7.4 programming environment on a 2.54 GHz Intel single-core processor computer with 512 MB of RAM. Our results are reported for the following four PRP algorithms:

AN1 : Algorithm 1;
AN2 : Algorithm 1 with the acceptance condition (33);
Max : Algorithm 1 with the acceptance condition (34);
GL : Algorithm 1 with with initial steplength s_k computed by (35).

It is observed that L, the Lipschitz constant, plays an important role in the algorithm 1, but L is not generally known in practical computation. So, we need to estimate it at each iteration for using in the proposed line search. we can set a large L_k to guarantee the global convergence. However, if L_k is very large then α_k will be very small and will slow the convergence rate of descent methods. On the other hand, very small values of L_k may fail to guarantee the global convergence. Thus, it is better to set an adequate estimation L_k at each iteration. Recently, some approaches for estimating L were proposed in [18, 19]. For example, Shi and Guo [19] proposed an approximation for the Lipschitz constant in the kth iteration as follows.

$$L_k = \max(L_0, \frac{\|g_k - g_{k-1}\|}{\|x_k - x_{k-1}\|}), \tag{36}$$

with $L_0 > 0$.

In all algorithms, we set $\delta = \gamma = 0.25$, $\rho = 0.9$, $c = 0.51$ and $L_0 = 3$. We also use (36) to estimate L for using in the proposed line search. All attempts to solve the test problems were limited to achieving a solution with

$$\|g_k\| \leq 10^{-6}\|g_0\|, \tag{37}$$

the numerical results are given in Table 1 which the columns have the following meanings:

Problem : The test problem name;
Dim : The dimension of the test problem;
$\mathbf{n_i}$: The total number of iterations;
$\mathbf{n_f}$: The total number of function evaluations.

In the following, we offer some observations about the numerical results related to the total number of iterations from Table 1. First, we observe that the methods AN1, AN2 and Max that use the same initial steplength are superior than the method GL that uses another steplength. When comparing the methods AN1, AN2 and Max, we see that the methods AN2 and Max are faster than the methods AN1.

At the same time, for more comprehensive comparison between the methods, we adopt the performance profiles of Dolan and Moré [7] to to evaluate the number of iterations and the number of function evaluations. In Figures 1 and 2, the vertical axis gives the fraction P of problems for which any given method is within a factor τ of the best performance. The left axis of the plot gives the percentage of the test problems for which a method needs least iterations. The right side of the plot gives the percentage of the test problems that are successfully solved by each of the methods. Clearly, the right side is a measure of an algorithms robustness.

Figures 1 and 2 give the performance problems of the four algorithms for the number of iterations and function evaluations, respectively. These figures show that in the perspective of the number of iterations AN1 and Max are more robust than AN2 and GL, but AN2 is the fastest method in more than 45% of the problems, in the perspective of the number of function evaluations, it is seen that from the perspective of robustness all of them are competitive with each other, but AN2 is still the fastest method in more than 45% of the problems. It is interesting to note that if AN2 method does not use the new initial

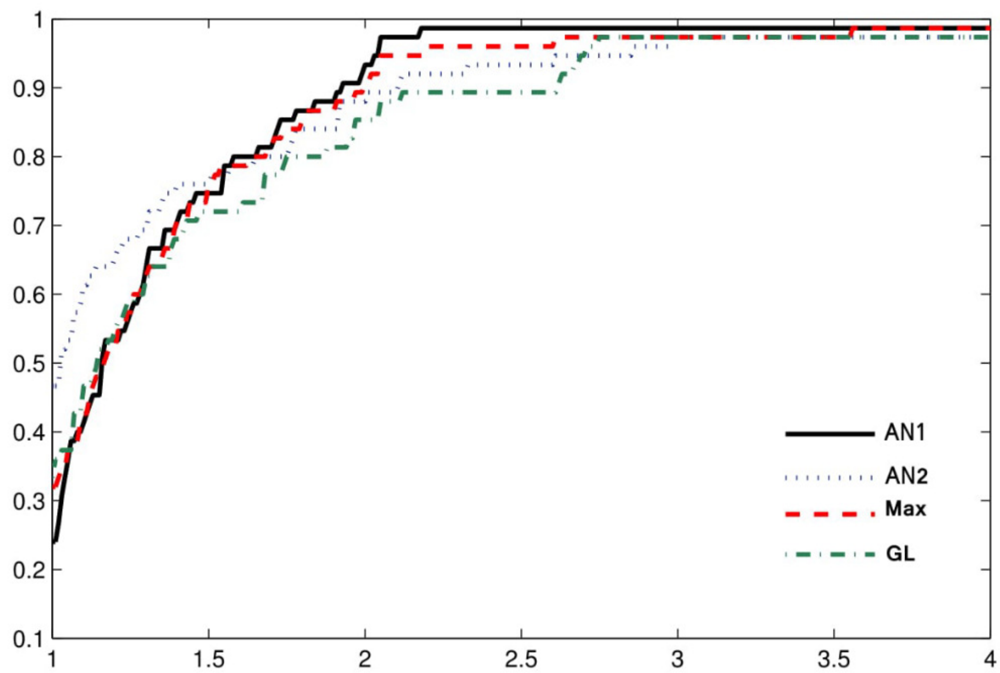

FIGURE 2. Function evaluation performance profiles for the four algorithms

steplength in line search procedure and uses the usual Armijo line search , then it will not be globally convergent Thus, the new initial steplength s_k used in line search procedure is effective in theory and practical .

Table 1: Numerical results.

Problem	Dim	AN1	AN2	Max	GL
		n_i/n_f	n_i/n_f	n_i/n_f	n_i/n_f
Almost.P. Quad.	1000	827/888	734/784	705/764	990/1048
ARGLINB	1000	8/378	7/367	8/367	23/392
ARGLINC	1000	8 / 377	7 / 366	7 / 374	23/390
ARWHEAD	1000	124/203	294/361	205/282	221/298
BDQRTIC	1000	728/ 878	674/803	741/894	909/1141
BG2	100	584 /632	996/997	996/997	382/383
Broyden Tridiag.	10000	138/151	277/289	178/216	207/214
COSINE	10000	26/27	26/27	26/27/1	73/74
CUBE	1000	581/643	587/645	519/580	1168/1224
Diagonal 1	1000	546/574	546/574	546/574	376/423
Diagonal 2	100	4785/4786	4785/4786	4785/4786	4799/4800

Continued on next page

Table 1 – continued from previous page

Problem	Dim	AN1	AN2	Max	GL
		n_i/n_f	n_i/n_f	n_i/n_f	n_i/n_f
Diagonal 3	1000	845/897	708/751	727/777	644/693
Diagonal 4	10000	251/306	259/283	226/278	210/265
Diagonal 5	1000	67/68	67/68	67/68	81/82
Diagonal 7	10000	8/9	8/9	8/9	23/24
Diagonal 8	10000	7/8	7/8	7/8	20/21
Diagonal 9	1000	5049/5148	5025/5119	4739/4820	2390/2677
DIXMAANA	6000	26/48	19/27	26/46	34/35
DIXMAANB	6000	24/41	18/26	24/40	37/51
DIXMAANC	6000	24/45	17/31	24/43	36/52
DIXMAAND	1500	23/44	17/38	24/44	35/50
DIXMAANE	6000	76/77	5278/5287	76/77	806/807
DIXMAANF	6000	2303/2311	1205/1213	2307/2314	2267/2269
DIXMAANG	6000	1503/1518	730/744	1053/1518	1510/1518
DIXMAANH	6000	520/557	431/452	533/569	533/570
DIXMAANJ	1500	2988/2992	2528/2536	2984/3021	2409/2411
DIXMAANL	6000	275/309	254 / 275	445/ 447	711/744
DQDRTIC	5000	177 / 222	84 / 121	174 /217	203/247
EDENSCH	1000	39 /56	52/ 65	41/ 57	45/53
ENGVAL1	1000	69/146	54/94	56/93	36/72
E. BD1	10000	34/35	34/35	34/35	39/40
E. Beale	1000	445/463	460/468	423/440	462/482
E. Cliff	10000	1/105	19/244	1/105	1/105
E.DENSCHNF	10000	25/76	21/66	26/75	32/79
E. Freud.Roth	1000	590/702	1289/1343	1348/1416	971/1215
E.Maratos	1000	805/848	1261/1327	977/1019	914/995
E.Penalty	10000	23/278	17/266	23/277	33/284
E.Powell	1000	1003/1050	1070/1115	989/1035	488/528
E. PSC1	10000	26/62	19/40	26/60	33/67
E. quad. exp. EP1	10000	7/59	7/46	7/57	23/74
E. quad. penalty QP1	10000	23/125	11/96	23/123	32/133
E. quad. penalty QP2	10000	18/117	146/231	18/115	29/126
E. TET	10000	33/35	31/36	33/35	41/43
E. Tridiag. 1	1000	3534/3535	3534/3535	3534/3535	3545/3546
E. Tridiag.2	10000	67/68	67/68	67/68	131/132
E. Wood	10000	596 / 659	1453/1515	490 / 552	454/510
Fletcher	100	4492 / 4548	4380 / 4415	4053 / 4106	2874/2935

Continued on next page

Table 1 – continued from previous page

Problem	Dim	AN1	AN2	Max	GL
		n_i/n_f	n_i/n_f	n_i/n_f	n_i/n_f
Full Hessian FH1	1000	927/1108	1053/1224	899/1078	1433/2115
Full Hessian FH3	1000	7/90	7/81	7/88	23/105
G. PSC1	10000	376/408	592/620	322/353	375/401
G.Quartic	10000	51/58	101/121	97/103	28/29
G. Rosenbrock	100	8681/8758	8895/ 8939	8746 / 8820	7375/7488
G. Tridiag. 1	10000	55/56	55/56	55/56	42/43
G. Tridiag. 2	10000	157/166	215/231	157/166	73/81
HARKERP2	1000	297 / 444	208 / 324	238 / 383	170/317
HIMELH	10000	30/31	30/31	30/31	34/35
LIARWHD	1000	2763/2859	2162/2231	3362/3445	1848/2044
NONDIA	10000	8 / 157	7 / 125	8 / 155	23/172
NONDQUAR	10000	148/250	150/242	170/270	177/278
NONSCOMP	1000	712/750	45/73	708/744	750/785
Par.Per.Quad.	1000	46/121	57/124	46/119	54/127
Per.Quad.tic	1000	720 / 783	847/897	646 / 706	585/645
Per. quad. diag.	5000	109 / 194	67 / 135	122 / 204	152/236
Per.Tridia. quad.	5000	1292/1383	1615/1680	1449/1538	1233/1322
POWER	100	5372/5445	5204/5275	5420/5491	2432/2501
Quadratic QF1	1000	768/813	726/769	874/917	596/637
Quadratic QF2	1000	1167/1217	1167/1215	1165/1213	832/901
QUARTC	1000	2846/2847	2846/2847	2846/2847	7643/7644
Raydan 1	1000	758/798	783/810	805/842	783/812
Raydan 2	1000	58/59	58/59	58/59	72/73
SINCOS	10000	26/62	19/40	26/60	33/67
Staircase 1	1000	1463/1583	1399/1509	799/917	2278/2397
Staircase 2	1000	1871/1991	1077/1187	2203/2321	1044/1163
TRIDIA	1000	7112/7180	7404/7454	7256/7322	4155/4217
Vardim	10000	23/606	16/594	23/605	33/610

6. Conclusion

It is well known that the sufficient descent condition is very important to the global convergence of the nonlinear conjugate gradient methods while the direction generated by a conjugate gradient method may not have sufficient descent condition, namely the PRP method. In this paper, we introduced a new Armijo-type line search algorithm based on the approach proposed by Grippo and Lucidi [10]. The proposed procedure is the same with condition (11) with a new procedure for computing the initial value of the steplength such that the direction generated by the PRP conjugate gradient method has sufficient descent property and ensures that the PRP conjugate gradient method AN1 is globally convergent for general functions under proper conditions. In addition algorithms AN2 and Max that use the new initial steplength are globally convergent and competitive with another. It is interesting to note that if AN2 method doesn't use the new initial steplength in line search procedure and uses the usual Armijo line search, then it will not be globally convergent. Thus, the new initial steplength s_k used in the new line search procedure is effective in theory and practical.

References

[1] Andrei, N., (2008), An Unconstrained Optimization Test Functions Collection, Advanced Modeling and Optimization, 10(1), 147–161.

[2] Dai, Y.H., (2001), New properties of a nonlinear conjugate gradient method, Numerische Mathematik, 89, 83–98.

[3] Dai, Y.H., (1997), Analysis of conjugate gradient methods. Ph.D. thesis, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences.

[4] Dai, Y.H., (2002), Conjugate gradient methods with Armijo-type line searches, Acta Mathematicae Applicatae Sinica (English Series), 18(1), 123–130.

[5] Dai, Y.H., Yuan, Y., (1996), Convergence properties of the Fletcher-Reeves method, IMA Journal of Numerical Analysis, 16, 155–164.

[6] De L.R., Gaudioso, M., Grippo, L., (1984), Stopping criteria for linesearch methods without derivatives, Mathematical Programming, 30, 285–300.

[7] Dolan, E., Moré, J.J., (2002), Benchmarking optimization software with performance profiles, Mathematical Programming, 91, 201–213.

[8] Fletcher, R. and Reeves, C., (1964), Function minimization by conjugate gradients, Computer Journal, 7, 149–154.

[9] Gilbert, J.C., Nocedal, J., (1992), Global convergence properties of conjugate gradient methods for optimization, SIAM Journal Optimization, 2, 21–42.

[10] Grippo, L., Lucidi, S., (1997), A globally convergent version of the Polak-Ribiére conjugate gradient method, Mathematical Programming, 78, 375–391.

[11] Hager, W.W., Zhang, H.C., (2005), A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM Journal Optimization, 16, 170–192.

[12] Hestenes, M.R., Stiefel, E., (1952), Method of conjugate gradient for solving linear systems, Journal of research of the National Bureau of Standards, 49, 409–436.

[13] Polak, E., Ribiére, G., (1969), Note sur la convergence de directions conjugées. Rev, Francaise Informat Recherche Opertionelle, 3e Année, 16, 35–43.

[14] Powell, M.J.D., (1984), Nonconvex minimization calculations and the conjugate gradient method, Numerical Analysis, Dundee, 1983, Lecture Notes in Mathematics, Vol. 1066, Springer-Verlag, Berlin. 122–141.

[15] Powell, M.J.D., (1986), Convergence properties of algorithms for nonliniear optimization, SIAM Review, 26, 487–500.

[16] Powell, M.J.D., (1977), Restart procedures of the conjugate gradient method, Mathematical Programming, 2, 241–254.

[17] Shanno, D.F., (1978), On the convergence of a new conjugate gradient algorithm. SIAM Journal Numerical Analysis, 15, 1247–1257.

[18] Shi, Z.J., Shen, J., (2005), Step-size estimation for unconstrained optimization methods, Journal of Computational and Applied Mathematics, 24, 399–416.

[19] Shi, Z.J., Guo, J., (2009), A new family of conjugate gradient methods, Journal of Computational and Applied Mathematics, 224, 444–457.

[20] Yuan, Y., (1993), Analysis on the conjugate gradient method, Optimization Methods and Software, 2, 19–29.

Hadi Nosratipour was born in Eslamabad-e Gharb, in Iran, in 1985. He obtained his PhD degree in Applied Mathematics in Operation Research area from Damghan University of Damghan, Iran in 2017. He obtained his MSc degree in Applied Mathematics in Operation Research area from Razi University of Kermanshah, Iran in 2011. His research interests include nonlinear optimization and optimal control.

Keyvan Amini is a Professor in Department of Mathematics, Razi University, Kermanshah, Iran. He was born in Kermanshah in Iran in 1974. He obtained his PhD degree in Applied Mathematics from Sharif University of Technology,Tehran, Iran in 2003. His area of interest includes nonlinear optimization, systems of nonlinear equations and Interior-point Methods.