

A globally convergent gradient-like method based on the Armijo line search

Ahmad Kamandi^{†*}, Keyvan Amini[‡]

[†]*Department of Mathematics, University of Science and Technology of Mazandaran, Behshahr, Iran*

[‡]*Department of Mathematics, Razi university, Kermanshah, Iran*
Email(s): ahmadkamandi@mazust.ac.ir, kamini@razi.ac.ir

Abstract. In this paper, a new conjugate gradient-like algorithm is proposed to solve unconstrained optimization problems. The step directions generated by the new algorithm satisfy sufficient descent condition independent of the line search. The global convergence of the new algorithm, with the Armijo backtracking line search, is proved. Numerical experiments indicate the efficiency and robustness of the new algorithm.

Keywords: Unconstrained optimization, conjugate gradient algorithm, global convergence, Armijo condition.

AMS Subject Classification 2020: 65K05 , 65K10.

1 Introduction

Motivated by numerous real-world applications such as machine learning [2, 3], big data [5], and compressive sensing [4, 22], it is essential to tackle large scale optimization problems and to design adapted efficient and robust algorithms, which are computationally tractable and globally convergent.

Consider the following unconstrained optimization problem:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in \mathbb{R}^n, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable function. Among numerical algorithms for solving (1), line search algorithms start from an initial guess, x_0 , for the solution of the problem

*Corresponding author.

and generate a sequence of iterations $\{x_k\}$ by the following recurrence relation:

$$x_{k+1} = x_k + \alpha_k d_k, \quad (2)$$

where α_k is the steplength and d_k is the search direction at iteration k . These algorithms try to find a new iterate x_{k+1} with a lower function value than x_k . To this aim, the direction d_k needed to be a downhill direction. The steepest descent method is a well-known line search method that uses the direction $d_k = -\nabla f(x_k)$, along which f decreases most rapidly, at every iteration.

Having the direction d_k , the ideal choice for the steplength α_k would be the global minimizer of

$$f(x_k + \alpha d_k), \quad \alpha > 0. \quad (3)$$

The procedure of finding exact minimizer of (3) is called exact line search and, in general, it is computationally expensive. More practical strategies perform an inexact line search to identify a steplength that achieves adequate reductions in f at minimal cost [20]. A popular inexact line search condition is

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k d_k^T g_k, \quad (4)$$

where $g_k = \nabla f(x_k)$ and $c_1 \in (0, 1)$. Inequality (4) is sometimes called the Armijo condition. An appropriate steplength α_k satisfying Armijo condition can be found by a so-called backtracking approach.

Another important procedure to find the steplength is known as the Wolfe line search for which the steplength α_k must satisfy the following conditions:

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + c_1 \alpha_k d_k^T g_k, \\ \nabla f(x_k + \alpha_k d_k)^T d_k &\geq c_2 d_k^T g_k, \end{aligned}$$

where $0 < c_1 < c_2 < 1$. Strong Wolfe conditions are also a modification of the Wolfe conditions that require α_k to satisfy

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + c_1 \alpha_k d_k^T g_k, \\ |\nabla f(x_k + \alpha_k d_k)^T d_k| &\leq c_2 |d_k^T g_k|. \end{aligned}$$

It is generally accepted that the steepest descent method is badly affected by ill-conditioning. The behaviour of this method is investigated comprehensively for a two dimensional case in [23]. Conjugate gradient (CG) methods are an important class of line search methods for solving (1); especially for the case that the dimension of the problem is large. The search direction in CG methods is determined by

$$d_k = \begin{cases} -g_0, & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \quad (5)$$

where β_k is a parameter that is called CG parameter and g_k is the gradient of the objective function at x_k .

The first nonlinear conjugate gradient method was introduced by Fletcher and Reeves in the 1960s [10]. The main difference among CG methods is in the formulas of computing their parameter for example:

$$\begin{aligned} \beta_k^{HS} &= \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} & [16], & & \beta_k^{FR} &= \frac{\|g_k\|^2}{\|g_{k-1}\|^2} & [10], \\ \beta_k^{PRP} &= \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} & [21], & & \beta_k^{CD} &= \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}} & [9], \\ \beta_k^{LS} &= \frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}} & [18], & & \beta_k^{DY} &= \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} & [6], \\ \beta_k^N &= (y_{k-1} - 2d_{k-1} \frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}})^T \frac{g_k}{d_{k-1}^T y_{k-1}} & [14]. \end{aligned}$$

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

When the objective function f is a strongly convex quadratic all of the above CG parameters are equivalent with an exact line search and the CG methods, obtained from them, converge to the global minimizer of f in at most n iterates. The global convergence properties of these and some other CG methods are reviewed in [15].

The Wolfe and the strong Wolfe conditions play an important role in establishing the global convergence of many CG methods [6, 11, 17, 19, 24]. But finding the steplength that satisfies these conditions, need some additional gradient evaluations. So, the Wolfe and the strong Wolfe line search are more expensive than the Armijo line search.

Another important property of the step direction, in convergence analysis of a CG method, is the ‘‘sufficient descent’’ condition. A direction d_k is called a sufficient descent direction if there exists a positive parameter c such that

$$d_k^T g_k \leq -c\|g_k\|^2. \tag{6}$$

The pioneer work about the global convergence of FR method with inexact line search was proposed by Al-Baali [1]. He proved that FR method generates sufficient descent directions and this method is globally convergent when the steplength satisfy the strong Wolfe conditions with $0 < c_1 < c_2 < \frac{1}{2}$. Liu et al. [13] and Dai and Yuan [6] extended this result to $c_2 = \frac{1}{2}$. It is shown that FR method with the strong Wolfe line search may not be a descent direction for the case that $c_2 > \frac{1}{2}$ [7].

For FR method, neither Armijo line search nor Wolfe line search, guarantee that the directions generated by this method are sufficient descent directions. In 2006 [25], Zhang et al. proposed a modified FR conjugate gradient method. In their method, the direction d_k is computed by

$$d_k = \begin{cases} -g_0, & \text{if } k = 0, \\ -\theta_k g_k + \beta_k^{FR} d_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{7}$$

where

$$\theta_k = \frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2}.$$

From the definition of d_k by (7), one can easily get $g_k^T d_k = -\|g_k\|^2$. Zhang et al., proved that the their method is globally convergent when the steplength, α_k , satisfies the following Armijo-type condition:

$$f(x_k + \alpha d_k) \leq f(x_k) + \sigma_1 \alpha d_k^T g_k - \sigma_2 \alpha^2 \|d_k\|^2, \tag{8}$$

where σ_1 and σ_2 are some positive constants.

In this paper, inspired by the work of Zhang et al. [25], we propose a new conjugate gradient-like method. The step directions generated by the new method are sufficient descent directions independent of the line search is used to compute the steplength. The new CG-like method is globally convergent under the Armijo condition. Numerical tests indicate the efficiency of the new method in solving a collection of unconstrained test problems from CUTEst package.

The rest of this paper is organized as follows. In the next section, the new CG-like method is introduced. Global convergence of the new method is analyzed in Section 3. Numerical results are reported in Section 4. Some conclusions are made in Section 5.

2 The new method

In this section, based on the pervious section discussion, we propose a new CG-like method. The sequence of iteration x_k in the new method is obtained from (2) for which the direction d_k is computed by (5). While the parameter β_k in the new method is computed by

$$\beta_k^{new} = \tau \frac{\|g_k\|}{\|d_{k-1}\|}, \quad (9)$$

where $\tau \in (0, 1)$. Note that, by $\tau = 0$ the new method is equivalent to the steepest descent method.

By the new parameter (9), we follow two goals: Firstly, to obtain a computationally inexpensive conjugate gradient like direction with sufficient descent property (6), in order to establish the well known Zoutendijk condition [26] when the steplength satisfies the Armijo condition (4). Secondly, to make the norm of the direction d_k , obtained from the new parameter, bounded by a fixed coefficient of $\|g_k\|$ in order to derive the global convergence of the new CG-like method.

Note that, for the direction d_k defined by (5), with the CG parameter computed by (9), we have

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

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

$$d_k^T g_k \leq -\|g_k\|^2 + \tau \|g_k\|^2 = -(1 - \tau) \|g_k\|^2. \quad (10)$$

So, the new direction d_k is a sufficient descent direction independent of the line search. For this direction, we also have

$$\|d_k\| \leq (1 + \tau) \|g_k\|. \quad (11)$$

Relations (10) and (11) illustrates two basic properties of the new CG-like direction which are provided by the new parameter (9). These properties will be used to prove the global convergence of the new method.

In the new CG-like method, the steplength α_k is determined such that satisfies the Armijo condition. To this aim, we use a backtracking approach to compute the steplength. Now, we are ready to propose the algorithm of the new CG-like method.

Algorithm 1 (New CG-like method)

- Step 1.** Given positive constants $c_1, \epsilon, \tau \in (0, 1)$. Choose an initial point $x_0 \in \mathbb{R}^n$, compute g_0 and set $k = 0$.
- Step 2.** If $\|g_k\| \leq \epsilon$ stop.
- Step 3.** If $k = 0$ set $d_k = -g_k$. Otherwise, compute the parameter β_k^{new} from (9) and the direction d_k from (5).
- Step 4.** Compute the steplength α_k that satisfies (4).
- Step 5.** Set $x_{k+1} = x_k + \alpha_k d_k$. Increase k by one, compute g_k and go to Step 2.
-

3 Convergence properties

In this section, we analyze the global convergence of the new CG-like method. To this aim, similar to [25], we made the following assumptions:

(H1) The objective function $f(x)$ has a lower bound on the level set

$$\mathcal{L} = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0), x_0 \in \mathbb{R}^n\}.$$

(H2) The objective function $f(x)$ is continuously differentiable and its gradient is Lipschitz continuous on a neighbourhood N of \mathcal{L} , namely, there exists a constant $L > 0$, such that

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

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

Lemma 1. *Let the steplength α_k be generated by Algorithm 1. Then, under the assumptions H1 and H2, there exists a positive constant C such that*

$$\alpha_k \geq C \frac{\|g_k\|^2}{\|d_k\|^2}. \tag{12}$$

Proof. There are two possible cases for the steplength α_k generated by Algorithm 1.

Case 1: $\alpha_k = \bar{\alpha}$. In this case, from (10), we have

$$\alpha_k \|d_k\| \|g_k\| \geq -\alpha_k d_k^T g_k = -\bar{\alpha} d_k^T g_k \geq \bar{\alpha}(1 - \tau) \|g_k\|^2.$$

So,

$$\alpha_k \geq \bar{\alpha}(1 - \tau) \frac{\|g_k\|}{\|d_k\|},$$

the square of this inequality along with $\alpha_k = \bar{\alpha}$ yield that (12) satisfies with $C = \bar{\alpha}(1 - \tau)^2$.

Case 2: $\alpha_k < \bar{\alpha}$. In this case, by the fact that $\rho^{-1}\alpha_k$ does not satisfy (4), we have

$$f(x_k + \rho^{-1}\alpha_k d_k) > f(x_k) + c_1 \rho^{-1}\alpha_k d_k^T g_k. \tag{13}$$

By the mean value theorem there exists a $t_k \in (0, 1)$, such that

$$\begin{aligned} f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k) &= \rho^{-1}\alpha_k \nabla f(x_k + t_k \rho^{-1}\alpha_k d_k)^T d_k \\ &= \rho^{-1}\alpha_k d_k^T g_k + \rho^{-1}\alpha_k [\nabla f(x_k + t_k \rho^{-1}\alpha_k d_k) - g_k]^T d_k. \end{aligned} \quad (14)$$

This equation, along with the assumption H2, result in

$$f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k) \leq \rho^{-1}\alpha_k d_k^T g_k + L\rho^{-2}\alpha_k^2 \|d_k\|^2. \quad (15)$$

From (10), (13) and (15) we can conclude that

$$\alpha_k > \frac{\rho(1-c_1)(1-\tau)\|g_k\|^2}{L\|d_k\|^2}.$$

This inequality means that (12) satisfies with $C = \frac{\rho(1-c_1)(1-\tau)}{L}$.

By the two above cases, the inequality (12) is always valid with

$$C = \min\{\bar{\alpha}(1-\tau)^2, \frac{\rho(1-c_1)(1-\tau)}{L}\}.$$

So, the proof is completed. \square

The next lemma is known as Zoutendijk condition [26].

Lemma 2. *Suppose that H1 and H2 hold and d_k is generated by Algorithm 1. Then*

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \quad (16)$$

Proof. From (4), for any $k \geq 0$, we have

$$\sum_{j=0}^k -\alpha_j d_j^T g_j \leq f(x_0) - f(x_{k+1}). \quad (17)$$

By the fact that the direction d_k is a sufficient descent direction and the Armijo condition is valid for each iteration, the sequence $\{f(x_k)\}$ is decreasing. So, the assumption H1 results that this sequence is convergent. Therefore, by taking limit from the inequality (17), when $k \rightarrow \infty$, we have

$$\sum_{j=0}^{\infty} -\alpha_j d_j^T g_j < \infty. \quad (18)$$

From the inequalities (18), (10) and (12), we have

$$\sum_{j=0}^{\infty} C(1-\tau) \frac{\|g_j\|^4}{\|d_j\|^2} < \infty. \quad (19)$$

So, the proof is completed. \square

Now, by the inequality (16) and (11), we can easily conclude that

$$\sum_{k=0}^{\infty} \|g_k\|^2 < \infty. \quad (20)$$

The inequality (20) results the global convergence of the new CG-like method.

Theorem 1. *Suppose that H1 and H2 hold and the sequence $\{x_k\}$ is generated by Algorithm 1, then $\lim_{k \rightarrow \infty} \|g_k\| = 0$.*

4 Numerical results

In this section, we report some numerical experiments that indicate the efficiency of Algorithm 1. To this aim, we implement Algorithm 1 (NEW), the conjugate gradient algorithm with the CG parameter proposed by Hager and Zhang [14] (HZ), Fletcher and Reeves (FR) algorithm and the modified Fletcher and Reeves (MFR) algorithm [25] in MATLAB environment on a laptop (CPU Corei7-2.5 GHz, RAM 12 GB) and compare their results obtained from solving a collection of 243 unconstrained optimization test problems from CUTEst collection [12]. The test problems and their dimension are listed in Table 1.

For all the considered algorithms, we set $\alpha_k = \max\{\bar{\alpha}\rho^i : i = 0, 1, 2, \dots\}$ which satisfies Armijo line search condition. For the initial steplength $\bar{\alpha}$, similar to [25], we set $\bar{\alpha} = s_{k-1}^T s_{k-1} / s_{k-1}^T y_{k-1}$, if $s_{k-1}^T y_{k-1} > 10^{-8}$ and $\bar{\alpha} = 1$, otherwise, where $s_{k-1} = g_k - g_{k-1}$. The other parameters are chosen as $\rho = 0.5$, $c_1 = 10^{-4}$ and $\tau = 0.002$.

Note that, in order to find an appropriate value for the parameter τ , we tested the numerical behavior of the new algorithm for different value of this parameter. Among $\{0, 0.001, 0.002, 0.003, 0.004, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, the best result is obtained for $\tau = 0.002$.

In our experiments the stopping tolerance for all of the algorithms is

$$\epsilon = 10^{-6} \|g_0\|.$$

Also, a failure is reported when, the total number of iterations exceeds 4000 (Case i) or the steplength α_k become less than $eps/10$ (Case ii). Where eps is the floating-point relative accuracy in MATLAB software. In the results of 243 problems 74, 6, 50 and 7 failures are reported for the FR, the New, the MFR and the HZ algorithms respectively. Where many of the failures in the FR are related to Case ii, while for the others the failures are related to Case i.

To visualize the whole behavior of the algorithms, we use the performance profiles proposed by Dolan and Moré [8]. The total number of function evaluations, the total number of iterations and the running time of each algorithm are considered as performance indexes. Note that at each iteration of the considered algorithms the gradient of the objective function is computed just one time, so the total number of iterations and the total number of the gradient evaluations are the same.

Fig. 1 illustrates the performance profile of the algorithms, where the performance index is the total number of function evaluations. It can be seen that the NEW is the best algorithm

Table 1: List of test problems.

Problem name	Dim	Problem name	Dim
ARGLINB	50, 100, 200	ARGLINC	50, 100, 200
BDQRTIC	100, 500, 1000, 5000	BROWNAL	100, 200, 1000
BRYBND	50, 100, 500	CHNROSNB	50
CHNRSNBM	50	EIGENALS	110
EIGENBLS	110	ERRINROS	50
ERRINRSM	50	EXTROSNB	100, 1000
FREUROTH	50, 100, 500, 1000, 5000	LIARWHD	100, 500, 1000, 5000
MANCINO	50, 100	MODBEALE	200, 2000
MSQRTALS	100	MSQRTBLS	100
NONDIA	50, 90, 100, 500, 1000, 5000	NONSCOMP	50, 100, 500, 1000, 5000
OSCIGRAD	100, 1000	OSCIPATH	100, 500
PENALTY1	50, 100, 500, 1000	PENALTY2	50, 100, 200
SPMSRTL	100, 499, 1000, 4999	SROSENBR	50, 100, 500, 1000, 5000
TQUARTIC	50, 100, 500, 1000, 5000	VAREIGVL	50, 100, 500, 1000, 5000
WOODS	100, 1000, 4000	ARWHEAD	100, 500, 1000, 5000
BOX	100	BOXPOWER	100, 1000
BROYDN7D	50, 100, 500, 1000	COSINE	100, 1000
CRAGGLVY	50, 100, 500, 1000, 5000	DIXMAANA	90, 300, 1500, 3000
DIXMAANC	90, 300, 1500, 3000	DIXMAAND	90, 300, 1500, 3000
DIXMAANE	90, 300, 1500, 3000	DIXMAANF	90, 300, 1500, 3000
DIXMAANG	90, 300, 1500, 3000	DIXMAANH	90, 300, 1500, 3000
DIXMAANI	90, 300, 1500, 3000	DIXMAANJ	90, 300, 1500, 3000
DIXMAANK	90, 300, 1500, 3000	DIXMAANL	90, 300, 1500, 3000
DIXMAANM	90, 300, 1500, 3000	DIXMAANN	90, 300, 1500, 3000
DIXMAANO	90, 300, 1500, 3000	DIXMAANP	90, 300, 1500, 3000
DQRTIC	50, 100, 500, 1000, 5000	EDENSCH	2000
ENGVAL1	50, 100, 1000, 5000	FLETCHCR	1000
FMINSURF	64, 121, 961, 1024	INDEFM	50
NCB20B	50, 1000, 2000	NONCVXU2	100, 1000, 5000
NONCVXUN	100, 1000, 5000	NONDQUAR	100, 1000, 5000
PENALTY3	50, 100	POWELLSG	60, 80, 100, 500, 1000, 5000
POWER	50, 75, 100, 500, 1000, 5000	QUARTC	100, 500, 1000, 5000
SCHMVETT	100, 500, 1000, 5000	SINQUAD	50, 100
SPARSINE	50, 100	SPARSQUR	50, 100, 1000, 5000
TOINTGSS	50, 100, 500, 1000, 5000	VARDIM	50, 100, 200
DIXON3DQ	100	DQDRTIC	50, 100, 500, 1000, 5000
HILBERTB	50	TESTQUAD	1000, 5000
TOINTQOR	50	TRIDIA	50, 100, 500, 1000, 5000

with probability around 60%, while the probability of solving a problem as the best algorithm are around 27%, 18% and 8% for the HZ, the MFR and the FR, respectively.

The performance index in Fig. 2 is the total number of iterations. From this figure we observe that the NEW obtains the most wins on approximately 55% of all test problems and the probability of being best algorithm is 30%, 25% and 8% for the HZ, the MFR and the FR

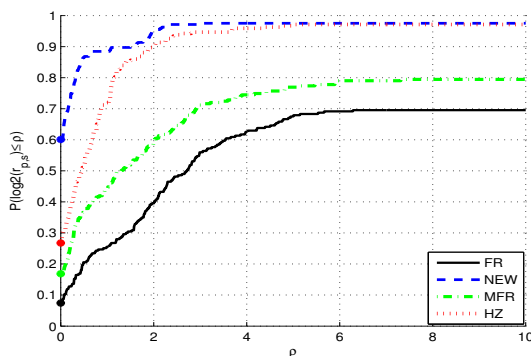


Figure 1: Performance profiles for the number of function evaluations.

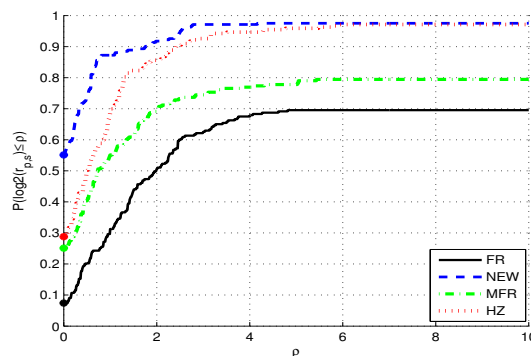


Figure 2: Performance profiles for the number of iterations.

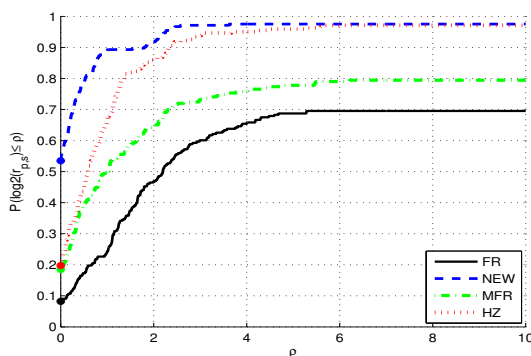


Figure 3: Performance profiles for the running times.

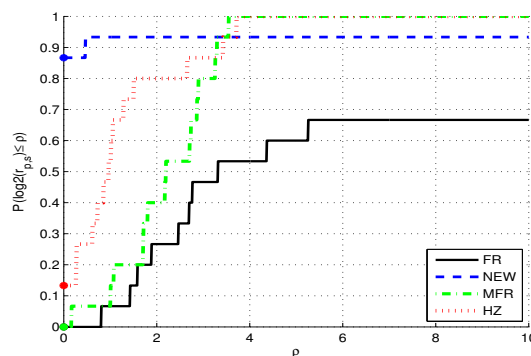


Figure 4: Performance profiles for the number of function evaluations (convex quadratic test problems).

respectively.

The performance profiles for the running times are illustrated in Fig. 3. From this figure, it can be observed that the NEW is the best algorithm. Another important factor of these three figures is that the graph of the NEW algorithm grows up faster than the other algorithms.

We also compare the mentioned algorithms in solving a set of convex quadratic problems from CUTEst package which are classified as QUR2. With the exact linesearch, as we expected, the FR and HZ methods perform the same and their results were much better than the NEW and the MFR. But, with the Armijo linesearch, the results were different.

The graphs of Fig. 4, Fig. 5 and Fig. 6, illustrate the performance of the four methods in solving 15 convex quadratic test problems which are listed in Table 2. These graphs indicate that the new methods is efficient than the other method in solving convex quadratic test problems, when the step length is computed by the Armijo line search.

From the presented results, we can conclude that the NEW algorithm is more efficient than the HZ, the MFR and the FR algorithms in solving unconstrained optimization problems.

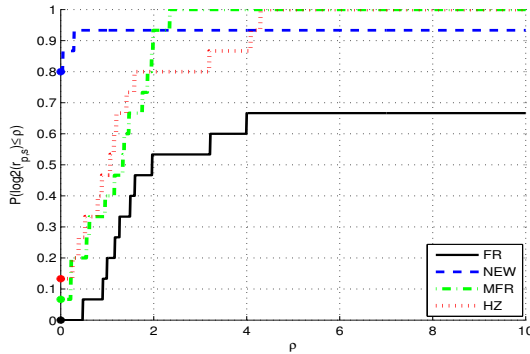


Figure 5: Performance profiles for the number of iterations (convex quadratic test problems).

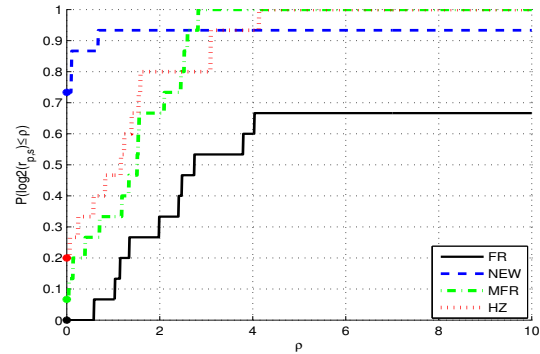


Figure 6: Performance profiles for the running times (convex quadratic test problems).

Table 2: List of convex quadratic test problems.

Problem name	Dim	Problem name	Dim
DIXON3DQ	100	TESTQUAD	1000, 5000
DQDRTIC	50, 100, 500, 1000, 5000	TOINTQOR	50
HILBERTB	50	TRIDIA	50, 100, 500, 1000, 5000

5 Conclusion

In this paper, we have proposed a new conjugate gradient-like algorithm. The step directions generated by the new algorithm satisfies sufficient descent condition independent of line search. The global convergence of the new algorithm with the Armijo backtracking line search has been investigated under some mild assumptions. Numerical experiments show the efficiency and robustness of the new algorithm for solving a collection of unconstrained optimization problems from CUTEst package.

References

- [1] M. Al-Baali, *Descent property and global convergence of the Fletcher-Reeves method with inexact line search*, IMA J. Numer. Anal. **5** (1985) 121–124.
- [2] L. Bottou, *Large-scale machine learning with stochastic gradient descent*, Proceedings of COMPSTAT 2010. Springer (2010) 177–186.
- [3] L. Bottou, F. Curtis, J. Nocedal, *Optimization methods for large-scale machine learning*, SIAM Rev. **60** (2018) 223–311.
- [4] E. Candes, *Compressive sampling*, Proceedings oh the International Congress of Mathematicians: Madrid, August 22–30, 2006: invited lectures. (2006) 1433–1452.

- [5] V. Cevher, S. Becker, M. Schmidt, *Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics*, IEEE Signal Process. Magz. **31** (2014) 32–43.
- [6] Y. Dai, Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J. Optim. **10** (1999) 177–182.
- [7] Y. H. Dai, Y. Yuan, *Nonlinear Conjugate Gradient Methods*, Shang Hai Scientific, Beijing, 2000.
- [8] E. Dolan, J. More, *Benchmarking optimization software with performance profiles*, Math. Program. **91** (2002) 201–213.
- [9] R. Fletcher, *Practical Methods of Optimization vol. 1: Unconstrained Optimization*, John Wiley and Sons, New York, 1987.
- [10] R. Fletcher, R. Colin, *Function minimization by conjugate gradients*, The computer journal **7** (1964) 149–154.
- [11] J.C. Gilbert, J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optim. **2** (1992) 21–42.
- [12] N. Gould, D. Orban, Ph. Toint, *CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization*, Comput. Optim. Appl. **60** (2015) 545–557.
- [13] L. Guanghui, H. Jiye, Y. Hongxia, *Global convergence of the Fletcher-Reeves algorithm with inexact linesearch*, Appl. Math. J. Chinese Univ. Ser, **10** (1995) 75–82.
- [14] W. Hager, H. Zhang, *A new conjugate gradient method with guaranteed descent and an efficient line search*, SIAM J. Optim. **16** (2005) 170–192.
- [15] W. Hager, H. Zhang, *A survey of nonlinear conjugate gradient methods*, Pac. J. Optim. **2** (2006) 35–58.
- [16] M. Hestenes, E. Stiefel, *Methods of conjugate gradients for solving linear systems*, Journal of Research of the NIST. **49** (1952) 409–436.
- [17] H. Liu, W. Haijun, X. Qian, F. Rao, *A conjugate gradient method with sufficient descent property*, Numer. Algorithms **70** (2015) 269–286.
- [18] Y. Liu, C. Storey, *Efficient generalized conjugate gradient algorithms Part 1: Theory*, J. Optim. Theory Appl. **69** (1991) 129–137.
- [19] I. Livieris, V. Tampakas, P. Pintelas, *A descent hybrid conjugate gradient method based on the memoryless BFGS update*, Numer. Algorithms **79** (2018) 1169–1185.
- [20] J. Nocedal, S. Wright, *Numerical Optimization*, Springer Science and Business Media, 2006.
- [21] E. Polakand, G. Ribiere, *Note sur la convergence de methodes de directions conjuguées*, Rev. Franais Informat. Recherche Operationnelle **3** (1969) 35–43.

- [22] J. Shen, S. Mousavi, *Least sparsity of p -norm based optimization problems with $p > 1$* , SIAM J. Optim. **28** (2018) 2721–2751.
- [23] J.R. Shewchuk, *An introduction to the conjugate gradient method without the agonizing pain*, Carnegie-Mellon University, Department of Computer Science 1994.
- [24] S. Yao, D. He, L. Shi, *An improved Perry conjugate gradient method with adaptive parameter choice*, Numer. Algorithms **78** (2018) 1255–1269.
- [25] L. Zhang, W. Zhou, D. Li, *Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search*, Numer. Math. **104** (2006) 561–572.
- [26] G. Zoutendijk, *Nonlinear programming, computational methods*, Integer and nonlinear programming (1970) 37–86.