Penerbit Akademia Baru

Journal of Advanced Research in Modelling and Simulations

Journal homepage: www.akademiabaru.com/armas.html
ISSN: XXXX-XXXX

A New Modified Three Term CG Search Direction for Solving Unconstrained Optimization Problems

Open Access

Kamilu Kamfa[1,3], M. Y. Waziri[2], Mustafa Mamat[1,*], Mohamad Afendee Mohamed[1], Puspa Liza Ghazali[4]

1 Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia
2 Department of Mathematical Science, Faculty of Science, Bayero University, Kano, Nigeria
3 Department of Mathematical Science, Faculty of Computing and Mathematics, Kano University of Science and Technology, Wudil, Nigeria
4 Faculty of Economics and Management Sciences, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia

ARTICLE INFO

Article history:
Received 18 March 2018
Received in revised form 6 February 2018
Accepted 8 February 2018
Available online 30 May 2018

ABSTRACT

The conjugate gradient (CG) method is a well know method for solving large scale unconstrained problem due to its simplicity and low memory requirement. For many years there are various modifications of this method, Zhang *et al*., [15] developed a three term CG method using a famous Polak-Ribiere-Polyak formula, This method satisfied the sufficient descent condition, but is not convergent under Wolfe condition. In this paper, a new three term CG method using a modified PRP formula is developed base on strong Wolfe condition. The new method satisfied the sufficient descent condition and is globally convergent. Numerical results generated using standard benchmark problem indicate that the proposed method performed better than the classical CG method.

Keywords:
Search direction, conjugate parameter, step length, global convergent, Wolfe line search

Copyright © 2018 PENERBIT AKADEMIA BARU - All rights reserved

1. Introduction

We consider an unconstrained optimization problem of the form

$$\min \{f(x): x \in R^n\} \quad (1)$$

where $f: R^n \to R^n$ is continuously diferentiable. Various problems of science, social science, economics and engineering can be cast into (1). There are various method for solving (1) such as conjugate gradient (CG), steepest method (SD) and Newton method (NM). Each method is design for a particular problem. Among all the methods CG method is often preferred, because of its simplicity and low memory requirement. This method generate a sequence of iterate $\{x_k\}$ using

* *Corresponding author.*
E-mail address: must@unisza.edu.my (Mustafa Mamat)

$$x_k = x_k + \alpha_k d_k \tag{2}$$

where $\alpha_k > 0$ is the step size obtained using a line search procedure and d_k is the search direction. CG method used the steepest descent direction at initial stage, whereas for the subsequent dierection it used

$$d_k = -g_k + \beta_k d_{k-1}, \ \ k \geq 1 \tag{3}$$

$g_k = \nabla f(x_k)$ is the gradient of the objective function at x_k and $\beta_k \in R$ is a scalar, known as conjugate parameter. There are Different β_k in the literature, each lead to a different CG method. Some of the well-known β_k include: the Hestenes–Stiefel (HS) [8], Fletcher–Reeves (FR) [9], Polak–Ribiere (PR) [10], Rivaie *et al*. (RMIL) [2] and Abashar et al. (AMRI) [1] developed the following update parameter respectively, defined as

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}, \qquad \beta_k^{FR} = \frac{\| g_k \|^2}{\| g_{k-1} \|^2}, \qquad \beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\| g_{k-1} \|^2}, \qquad \beta_k^{RMIL} = \frac{g_k^T(g_k - g_{k-1})}{\| d_k \|^2}$$

CG methods with FR formula converge globally but has poor numerical performance due to the jamming behaviour encounter along the iteration process. The PRP and HS methods has good numerical performance but they are not always convergent [6]. For many years, various researchers like Zoutendijk [4], Al-Baali [8] and Gilbert and Nocedal [7] have studies the global convergence properties of the FR, PRP and HS methods. To establish the convergence results of these methods, it is usually required that the step length should satisfy the inexact line search such as strong Wolfe, Armijo or standard Wolfe line search. Other line search procedure are some times use for example [9] use the exact line search to prove the convergence of RMIL method.

The Wolfe line search is define as

$$f(x_k + \alpha_k d_k) \ \leq f(x_k) + \mu \alpha_k \nabla f(x_k)^T d_k$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \ \leq -\sigma \nabla f(x_k) d_k^T \tag{4}$$

with $0 < \mu < \sigma < 1$. Whereas, the strong Wolfe line search is obtained using

$$f(x_k + \alpha_k d_k) \ \leq f(x_k) + \mu \alpha_k \nabla f(x_k)^T d_k$$

$$|\nabla f(x_k + \alpha_k d_k)^T d_k| \ \leq -\sigma \nabla f(x_k) d_k^T \tag{5}$$

To improved the performance of the classical _k, researchers like Andrei [8], Asrul *et al.*, [10] and Abdelrahman *et al.*, [14] have proposed a robust hybrid CG method that satisfies the global convergence properties. The class of those methods are called two term CG method. Recently, researchers focus on improving these class of methods with three term CG method. Preliminary results have shown that, the three term CG methods performance depend on how the conjugate parameters is being selected. Details of those results can be found in [4,6,7].

Yanlin Wu [16] develop a new three term CG method, namely NTT-PRP-CG-A. This method perform well compared to the famous one term or two terms CG methods. Also, is globally convergent under standard Wolfe line search method. The performance is basically associated with a restart property in the numerator. This property is what makes the PRP, HS and RMIL method efficient and reliable. Wei *et al*., [12] improved this properties by introducing a new parameter called VPRP $\frac{\|g_k\|}{\|g_{k-1}\|}$.

This parameter satisfies $\{*\}$ property [7] and has reduced the difficulties in proving the sufficient and global convergent properties of PRP method. Many researchers, adopt this new concepts, to proposed various CG method [3,12], the performance of those methods are quit impressive independent of any line search.

The rest of this paper is organised as follows. In section 2, we present the motivation and our new three term formula. In section 3 we gives the convergent analysis of the new method. the numerical results is presented in section 4 follows by the conclusion remark in section 5.

2. Motivation and New Formula

Wei *et al*., [12] developed a new CG method called VPRP

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

The numerator in this formula, plays an important role in making the method efficient and reliable. Yao *et al.,* [12] based on HS formula improved (10) as follows

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

Abashar *et al*., [3] proposed another modification of (10) by changing the denominator, as follows:

$$\beta_k^{AMRI} = \frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}\right)}{\|d_{k-1}\|^2}$$

This method also satisfies the sufficient descent property, independent of any line search and is globally convergent.

Recently, [16] developed a new three CG method under standard Wolfe line search namely NTT-PRP-CG-A. with the following search direction

$$d_k = -g_k + \frac{g_k^T \delta_k d_k - d_k^T g_k \delta_k}{\gamma_1 \|g_k\|^2 + \gamma_2 \|d_k\|\|\delta_k\| + \gamma_3 \|d_k\|\|g_k\|} \tag{7}$$

The different between the Zhang three term CG and [11] is the second and third terms in the denominator. Note that $g_k - g_{k-1}$ is kept so that the search direction would restart when ever the step taken is very small.

Motivated by this, we develop a new modified three term CG method with the following search direction.

$$d_k = \begin{cases} -g_k + \frac{g_k^T \delta_k d_{k-1} - d_{k-1}^T g_k \delta_k}{\mu_1 \|g_{k-1}\|^2 + 2\mu_2 \|d_{k-1}\|\|\delta_k\| + \mu_1 \|g_{k-1}\|\|d_{k-1}\| + |d_{k-1}^T g_k|} & k \geq 1 \\ -g_k & k = 0 \end{cases} \tag{8}$$

where $\delta_k = g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}$ is used to remedie any possible infinitely cycle of the PRP method and its variant as mentioned by [15]. Also, the denominator is different from the (11), this is a tiny alteration, that will guarantee another encouraging results for (12) and propel the global convergence for Wolfe conditions. Below we present the complete Algorithm, namely KMM6 (based on the author initial)

Algorithm 2.1; (KMM6)

Initialization, Given a starting point x_0, let $\varepsilon \in (0,1)$ for k = 0

Step 1 Terminate if $\|g_k\| < 10^{-6}$ or $k \geq 1000$

Step 2 Find the search direction using (8),

Step 3 Calculate the step size using Wolfe line search (4)

Step 4 Updated x_k using (2)

Step 5 Set k = k + 1, Go to Step 1.

3. Converegence Analysis of KMM6 Method

For an algorithm to convergence it must satisfy the su_cient descent condition and the global convergence properties under Wolfe line search. All the proof will be supported with numerical result generated using standard benchmark problems, we begin by making the following assumption.

Assumption 1

(i) The level set $\Omega = \{x \in R^n \mid f(x) \leq f(x_0)\}$ is bounded.

(ii) In some neighborhood N of Ω f is continuously differentiable and its gradient $g(x)$ is Lipschitz continuous, i.e. $\exists\ L > 0$ such that

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

To begin we show that the proposed search direction is sufficiently descent

Lemma 3.1. The search direction d_k defined by (8) satisfies the following condition

$$g_k^T d_k \leq -c \| g_k \|^2 \tag{9}$$

Where $c > 0$ is a constant, and

$$\|d_k\| \leq -c\|g_k\| \tag{10}$$

For $k \geq 0$,

Proof: for $k = 0$ it is obvious $g_0^T d_0 = -g_0^T g_0 = -\|g_0\|^2$ and $\|d_0\| = \|-g_0\| = \|g_0\|$ which means (9) is true with $c = 1$, now we show for $k \geq 1$ also (9) holds. From (8), we have

$$d_k = -g_k + \frac{g_k^T \delta_k d_{k-1} - d_{k-1}^T g_k \delta_k}{\mu_1 \|g_{k-1}\|^2 + 2\mu_2 \|d_{k-1}\| \|\delta_k\| + \mu_1 \|g_{k-1}\| \|d_{k-1}\| + |d_{k-1}^T g_k|}$$

multiply both side by g_k^T we have,

$$g_k^T d_k = -g_k^T g_k + g_k^T \left(\frac{g_k^T \delta_k d_{k-1} - d_{k-1}^T g_k \delta_k}{\mu_1 \|g_{k-1}\|^2 + 2\mu_2 \|d_{k-1}\| \|\delta_k\| + \mu_1 \|g_{k-1}\| \|d_{k-1}\| + |d_{k-1}^T g_k|} \right)$$

$$= -\|g_k\|^2 + g_k^T \left(\frac{g_k^T \delta_k d_{k-1} - d_{k-1}^T g_k \delta_k}{\mu_1 \|g_{k-1}\|^2 + 2\mu_2 \|d_{k-1}\| \|\delta_k\| + \mu_1 \|g_{k-1}\| \|d_{k-1}\| + |d_{k-1}^T g_k|} \right)$$

$$= -\|g_k\|^2$$

this means (9) is true independent of any line search. By (9) again , we can also get $\|g_k\| \leq \|d_k\|$ as follows. Taking the absolute value of both side of (9), we have

$$|g_k^T d_k| = |-\|g_k\|^2| = \|g_k\|^2$$

by Cauchy Schwartz inequality, we have

$$|g_k^T d_k| \leq \|g_k\| \|d_k\|$$

Therefore

$$\|g_k\|^2 \leq \|g_k\| \|d_k\|$$

$$\|d_k\| \leq \|g_k\| \quad (11)$$

Lemma 3.2. Let Assumption 1 and 2 hold, supposed x_0 is the initial point. Now consider any method in form of (2), in which d_k is a descent direction and α_k satisfies (4) or (5) respectively. Then

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

The above equation is refer as Zoutendijk condition. Normally it is consider when proving the global convergent of a new CG method. Further more this equation is equivalent to the following condition.

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} = +\infty \quad (12)$$

we now present the global convergent of Algorithm (2.1).

Lemma 3.3. Suppose that Assumption 1 is true and KMM6 generates the sequence $\{x_k, \alpha_k, d_k, g_k\}$. Then there exist a constant $\lambda > 0$ such that

$$\alpha_k \geq \lambda \quad (13)$$

$\forall\, k \geq 1$

Proof. Based on the Lipschitz condition and first inequalities of Wolfe line search (4) we get,

$$L\alpha_k = (g_k - g_{k-1})d_{k-1}$$
$$\geq -(1-\sigma)g_k^T d_k$$
$$\geq (1-\sigma)\|g_k\|^2$$

where the last equality comes from (8). using (9), we have

$$\alpha_k = \frac{(1-\sigma\|g_k\|^2)}{L\|d_k\|^2} \geq \frac{(1-\sigma\|g_k\|^2)}{L\gamma\|g_k\|^2} = \frac{(\sigma-1)}{L\gamma}$$

Now let $\lambda \in \left(0, \frac{1-\sigma}{L\gamma}\right)$ then the proof is complete.
Lemma 3.3 indicate that, the step length α_k has the lower band. This is an important part in proving the global convergent of KMM6 method.

Theorem 3.1. Let the condition of Lemma (3.2) and (3.3) hold and the sequence $\{x_k, \alpha_k, d_k, g_k\}$. Be generated by Algorithm 2.1, then

$$\lim_{k\to\infty} \|g_k\| = 0 \tag{14}$$

Proof
from (8) and lemma (3.3) we have,

$$d_k = \left\| -g_k + \frac{g_k^T \delta_k d_{k-1} - d_{k-1}^T g_k \delta_k}{\mu_1\|g_{k-1}\|^2 + 2\mu_2\|d_{k-1}\|\|\delta_k\| + \mu_1\|g_{k-1}\|\|d_{k-1}\| + |d_{k-1}^T g_k|} \right\|$$
$$\leq \|g_k\| + \frac{\|g_k^T \delta_k d_{k-1} - d_{k-1}^T g_k \delta_k\|}{\mu_1\|g_{k-1}\|^2 + 2\mu_2\|d_{k-1}\|\|\delta_k\| + \mu_1\|g_{k-1}\|\|d_{k-1}\| + |d_{k-1}^T g_k|}$$
$$\leq \|g_k\| + \frac{\|g_k^T \delta_k d_{k-1} - d_{k-1}^T g_k \delta_k\|}{\mu_1\|g_{k-1}\|^2 + 2\mu_2\|d_{k-1}\|\|\delta_k\| + \mu_1\|g_{k-1}\|\|d_{k-1}\| + |d_{k-1}^T g_k|}$$
$$\leq \|g_k\| + \frac{\|g_k\|\|\delta_k\|\|d_{k-1}\| + \|g_k\|\|\delta_k\|\|d_{k-1}\|}{\mu_1\|g_{k-1}\|^2 + 2\mu_2\|d_{k-1}\|\|\delta_k\| + \mu_1\|g_{k-1}\|\|d_{k-1}\| + |d_{k-1}^T g_k|}$$
$$\leq \|g_k\| + \frac{2\|g_k\|\|\delta_k\|\|d_{k-1}\|}{2\mu_2\|d_{k-1}\|\|\delta_k\|}$$
$$= \left(1 + \frac{1}{\mu}\right)\|g_k\| \tag{15}$$

Now letting $\sqrt{D} = \left(1 + \frac{1}{\mu}\right)$ we have $\|d_k\|^2 \leq D\|g_k\|^2$

$$\frac{1}{\|d_k\|^2} \geq \frac{1}{D\|g_k\|^2}$$
$$\frac{D\|g_k\|^4}{\|d_k\|^2} \geq \frac{\|g_k\|^4}{\|g_k\|^2} = \|g_k\|^2 \tag{16}$$

Therefore by (12), we have

$$\lim_{k\to\infty} \|g_k\|^2 \leq D \lim_{k\to\infty} \frac{\|g_k\|^4}{\|d_k\|^2} = 0 \tag{17}$$

Hence, the proof is completed.

4. Numerical Result

In this section, we present the numerical performance of KMM6 when compared with FR,MPRP, and AMRI method. The MPRP is the three term method developed by Zhang [15]. All the algorithm are coded in MATLAB R2013b and tested for some well know benchmark problem with $\|g_k\| < 10^{-6}$ as stopping condition $\mu_1 = \mu_2 = \mu_3 = 0.1, c_1 = 0.001$ and $c_2 = 0.86$. The results is then analysed using the performance profiles introduced by Dolan and More [2]. This profile compare and evaluate the performance of the set solver S on test P. ie, the probability of success, assuming that n_s and P_s problem exist. Details of the performance profile can be found in [5, 9].

Figure 1 and 2 shows that KMM6 method is e_cient with a good numerical result, that iswhy its curves appear at the top and reach 1. FR method has a good convergent rate, but it is numerically poor. Hence its curves appear below with 0.76 success. MPRP and AMRI methods fall between the two groups with 0.82 and 0.8 success respectively. Even-though, MPRP method performs well, but it is not robust enough to solve all the test problem.

Table 1
A list of the test problems

Test Function	Dimension	Source
Booth	2	Rivaie [2]
Nonscomp	2,4,10,100	Andrei [26]
Generalized Tridiagonal 1	100,500,1000,5000,20000,30000	Andrei [26]
Quadratic QF2	2,4,10,100,500	Andrei [26]
Diagonal 4	100,500,1000,5000,20000,30000	Andrei [26]
Extended Maratos	2,4,100,500,1000	Rivaie [2]
Extended Rosenbrock	100,500,1000,5000,20000,30000	Andrei [26]
Extended Himmeblau	100,500,1000,5000,20000,30000	Rivaie [2]
Freudenstein and Roth	100,500,1000,5000,20000,30000	Andrei [26]
Extended Beale	100,500,1000,5000,20000,30000	Rivaie [2]

5. Conclusion

In this paper, we present a new algorithm for solving unconstrained optimization problems. It is a three-term method that utilized the modi_ed PRP formula under Wolfe line search. The idea is to improved the performance of MPRP under Wolfe line search. Various standard benchmark test problems were used to demonstrate the e_ciency of our new method. The results show that KMM6 is robust, reliable and e_ective and can be used in place of the popular CG method.

References

[1] Andrei, Neculai. "An unconstrained optimization test functions collection." *Adv. Model. Optim* 10, no. 1 (2008): 147-161.

[2] Dolan, Elizabeth D., and Jorge J. Moré. "Benchmarking optimization software with performance profiles." *Mathematical programming* 91, no. 2 (2002): 201-213.

[3] Abashar, Abdelrahman, Mustafa Mamat, Mohd Rivaie, and Ismail Mohd. "Global convergence properties of a new class of conjugate gradient method for unconstrained optimization." (2017).

[4] Powell, Michael JD. "Nonconvex minimization calculations and the conjugate gradient method." In *Numerical analysis*, pp. 122-141. Springer, Berlin, Heidelberg, 1984.

[5] Rivaie, M., M. Mustafa, M. Ismail, and M. Fauzi. "A comparative study of conjugate gradient coefficient for unconstrained optimization." *Aus. J. Bas. Appl. Sci* 5 (2011): 947-951.

[6] Hager, William W., and Hongchao Zhang. "A new conjugate gradient method with guaranteed descent and an efficient line search." *SIAM Journal on optimization* 16, no. 1 (2005): 170-192.

[7] Gilbert, Jean Charles, and Jorge Nocedal. "Global convergence properties of conjugate gradient methods for optimization." *SIAM Journal on optimization* 2, no. 1 (1992): 21-42.

[8] Al-Baali, Mehiddin. "Descent property and global convergence of the Fletcher—Reeves method with inexact line search." *IMA Journal of Numerical Analysis* 5, no. 1 (1985): 121-124.

[9] Rivaie, Mohd, Mustafa Mamat, W. J. Leong, and M. Ismail. "A New conjugate gradient coefficient for large scale nonlinear unconstrained optimization." *Int. Journal of Math. Analysis* 6, no. 23 (2012): 1131-1146.

[10] Ibrahim, Mohd Asrul Hery, Mustafa Mamat, and Wah June Leong. "The hybrid BFGS-CG method in solving unconstrained optimization problems." In *Abstract and Applied Analysis*, vol. 2014. Hindawi, 2014.

[11] Kamfa, Kamilu Uba, Mustafa Mamat, Abdelrhaman Abashar, Mohd Rivaie, Puspa Liza Binti Ghazali, and Zabidin Salleh. "Another modified conjugate gradient coefficient with global convergence properties." (2017).

[12] Dai, Zhifeng, and Fenghua Wen. "Another improved Wei–Yao–Liu nonlinear conjugate gradient method with sufficient descent property." *Applied Mathematics and Computation* 218, no. 14 (2012): 7421-7430.

[13] Abashar, Abdelrhaman, Mustafa Mamat, Mohd Rivaie, and Zabidin Salleh. "Amodified dprp conjugate gradent method for unconstrained optimazation." (2017).

[14] Abdelrahman, Awad, Mustafa Mamat, Zabidin Salleh, and Mohd Rivaie. "A MODIFIED CONJUGATE GRADIENT METHOD WITH SUFFICIENT CONDITION AND CONJUGACY CONDITION." *Far East Journal of Mathematical Sciences* 99, no. 4 (2016): 525.

[15] Zhang, Li, Weijun Zhou, and Dong-Hui Li. "A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence." *IMA Journal of Numerical Analysis* 26, no. 4 (2006): 629-640.

[16] Wu, Yanlin. "A modified three-term PRP conjugate gradient algorithm for optimization models." *Journal of inequalities and applications* 2017, no. 1 (2017): 97.