Malaysian Journal of Mathematical Sciences 10(S) February: 91–101 (2016)
Special Issue: The 3rd International Conference on Mathematical Applications in Engineering 2014 (ICMAE'14)

MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES

Journal homepage: http://einspem.upm.edu.my/journal

A Modified NPRP Nonlinear Conjugate Gradient Method with Global Convergence Properties

Alhawarat, A.*[1], Mustafa, M.[2], Ismail, M.[1], and Zabidin, S.[1]

[1]*School of Informatics and Applied Mathematics, University Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia.*
[2]*Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia.*

*E-mail: abadee2010@yahoo.com**

ABSTRACT

Conjugate gradient method (CG), is one of many methods used to solve large scale optimization problems in many fields. In fact memory storage and speed of the algorithm to reach the optimum solution for general nonlinear unconstrained optimization functions are our goals. In this paper, we modified a new efficient method passes the sufficient decent condition, and it has global convergence under exact line search and strong Wolfe line search. Furthermore the numerical results to find the optimum solution of some test functions show the new modification has the best results in CPU time, and the number of iteration when it comparing with other conventional methods.

Keywords: Gradient Method, Exact Line Search, Strong Wolf Line Search, Global Convergence

1. Introduction

To find a solution for unconstrained optimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and its gradient is denoted by $g(x) = \nabla f(x)$, the CG method is one of the best methods designed to find sequence of points $\{x_k\}$ starting from initial point $\{x \in \mathbb{R}^n\}$ by using the iterative formula,

$$x_{k+1} = x_k + \alpha_k d_k,\ k = 0, 1, 2, 3, \ldots, \tag{2}$$

where x_k is the current iteration point and $\alpha_k > 0$ is the step size obtained by some line search. The search direction d_k is defined by;

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

where $g_k = g(x_k)$ and β_k is a scalar known as the CG coefficient of CG method. The most well known formulas for CG Methods are as follows;
$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$, (Fletcher and Reeves, 1964).
$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}$, (Polak and Ribiere, 1969).
In order to find the step length (α_k), there are many line searches are often used such as,
- Exact line search,

$$f(x_k + \alpha_k d_k) = \min f(x_k + \alpha d_k),\ \alpha > 0. \tag{4}$$

- Strong Wolf Powell (SWP) line search,

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \tag{5}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T d_k|, \tag{6}$$

where $\delta \in (0, \frac{1}{2})$ and $\sigma \in (0, 1)$ are two constants.
- Weak Wolfe- Powell line search,

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \tag{7}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k. \tag{8}$$

Recently,(Wei et al., 2006)
gave a new method which is a variant of PRP method that is,

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

It seems like original PRP method, which has been studied in both exact line search and inexact line search, and many modifications have appeared, as following,(Dai and Hong, 2011)
, and (Zhang, 2009)

$$\beta_k^{VHS} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T (g_k - g_{k-1})}, \; \beta_k^{NPRP} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} |g_k^T g_{k-1}|}{\|g_{k-1}\|^2},$$

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

For more, we suggest the reader to see the following references,(Hager and Zhang, 2006)
,(Dai 2011), (Al-Baali, 1985)
,(Rivaie, 2012)
, (Mamat and Mustafa, 2010)
,(Alhawarat, 2014)
, and (Ibrahim and Leong, 2014)
In this paper, we will present the new formula and the algorithm in Section 2. In addition the sufficient decent condition and the global convergence for the new method under exact line search and strong Wolf line search will be presented in Section 3. Finally we will discuss the numerical results and conclusion in Sections 4 and five respectively.

2. The new formula

In this section, we present our new β_k^{AMIPRP}, where $AMIPRP$ denotes to Ahmad, Mustafa, and Isamil, which is extended from β_k^{NPRP} method that is,

$$\beta_k^{AMIPRP} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} |g_k^T g_{k-1}|}{\|d_{k-1}\|^2}, \tag{9}$$

where $\|.\|$ means the Euclidean norm. The algorithm is given as follows:
Step 1: Initialization. Given x_0 set $k = 0$.
Step 2: Compute β_k based on Eq.(9).
Step 3: Compute d_k based on Eq.(3), if $\|g_k\| = 0$, then stop.
Step 4: Compute α_k based by some line search.
Step 5: Updating new point based on Eq.(2).
Step 6: Convergent test and stopping criteria, if $f(x_k) < f(x_{k+1})$ and $\|g_k\| \leq \epsilon$, then stop, otherwise go to Step 1 with $k = k + 1$.

3. Global Convergence analysis by several line searches for β_k^{AMIPRP} method

In this section, the convergent properties of β_k^{AMIPRP} be studied for exact an inexact line searches.

3.1 Descent condition

In exact line search the decent condition,

$$g_k^T d_k \leq -\|g_k\|^2 \text{ for } k \geq 0. \tag{10}$$

The following theorem establishes that β_k^{AMIPRP} achieves the decent condition under exact line search.

Theorem 3.1. *Consider a CG method with the search direction Eq.(4) and β_k^{AMIPRP} given as in Eq.(9), the Eq.(10) holds for all $k \geq 0$.*

Proof. From Eq.(3) if $k = 0$, then $g_0^T d_0 = -\|g_0\|^2$. For $k \geq 1$, multiply Eq.(3) by g_{k+1}^T, then

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

For exact line search easy to know $g_{k+1}^T d_k = 0$. Thus $g_{k+1}^T d_{k+1=-\|g_{k+1}\|^2}$. The proof is completed. □

To satisfy the global convergence properties for new method the following simplifications are needed,

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

By using Cauchy-Schwartz inequality:

$$\beta_k^{AMIPRP} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}|g_k^T g_{k-1}|}{\|d_{k-1}\|^2} \geq \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}(|g_{k-1}\|.\|g_k\|)}{\|d_{k-1}\|^2} \geq 0.$$

Thus we get,

$$\beta_k^{AMIPRP} \geq 0. \tag{11}$$

$$\beta_k^{AMIPRP} \leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2}. \tag{12}$$

3.2 Global convergence by exact line search

Assumption 1:
1. $f(x)$ is bounded from below on the level set $\Omega = \{x \in \mathbb{R}^n | f(x) \leq f(x_1)\}$, where x_1 is the starting point.
2. In some neighborhood N of Ω, f is continues and differentiable, and its gradient is Lipchitz continues, that is for any $x, y \in N$ there exists a constant $L \geq 0$ such that,

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

Lemma 3.1. *(Zoutendijk, 1970)*
. Suppose Assumption 1 is true. Consider any form of Eq.(3) for all k, and α_k satisfied one of the following line searches, exact line search, Armijo-Goldstein line search, or Wolfe-Powell line search. Then the following condition

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty, \tag{13}$$

holds.

Substitute Eq.(3) in Eq.(13) then it is equivalent to the following,

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{14}$$

Theorem 3.2. *Suppose that Assumptions 1 hold, and the sequence $\{x_k\}$ is generated by the Algorithm Eq.(2), and α_k is determined by Eq.(4), if $\|x_{k+1} - x_k\| \to 0$ while $k \to \infty$, then*

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

Proof. Suppose that there exist a constant $\epsilon > 0$ such that,

$$\|g_k\| \geq \epsilon \text{ holds for all } k \geq 0. \tag{15}$$

Consider θ_k is the angle between d_k and the steepest descent search direction $-g_k$, where

$$\cos\theta_k = -\frac{g_k^T}{\|g_k\|.\|d_k\|}. \tag{16}$$

By using Eq.(3) and Eq.(16), we indicate the following relation,

$$\|d_k\| = \sec\theta_k \|g_k\| \tag{17}$$

and

$$\beta_{k+1}\|d_k\| = \tan\theta_{k+1}\|g_{k+1}\|. \tag{18}$$

Combining Eq.(17) and Eq.(18), indicates

$$\tan\theta_{k+1} = \frac{\|g_{k+1}\|^2}{\|d_k\|^2}\sec\theta_k.\frac{\|g_k\|^2}{\|g_{k+1}\|}.$$

Since $\|g_k\|.\|g_{k+1}\| \leq \|g_k\|^2$, we have

$$\tan\theta_{k+1} \leq \frac{\|g_k\|^2}{\|d_k\|^2}\sec\theta_k \leq \frac{1}{\sec\theta_k} = \cos\theta_k.$$

Therefore, the angle between d_k and the steepest descent direction $-g_k$ is bounded away from $\frac{\pi}{2}$. So from Eq.(13), Eq.(14) and Eq.(15)

$$\sum_{k=0}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2} = \sum_{k=0}^{\infty}\|g_k\|^2(\cos\theta_k)^2 = \infty.$$

Which is contradiction with Eq.(14), so $\lim_{k\to\infty}\inf\|g_k\| = 0$. The proof is completed. □

3.3 Global convergence of β_k^{AMIPRP} method under SWP line search

In the following theorem, we discuss the sufficient decent condition, $g_k^T d_k \leq -c\|g_k\|^2$, where $k \geq 0$ and $c \in (0, 1)$ under SWP line search.

Theorem 3.3. *If the sequences g_k and d_k are generated by the methods Eq.(2), Eq.(3) and Eq.(13) with step length α_l is determined by Eq.(5) and Eq.(6) if $\sigma \in (0, \frac{1}{2}]$, then the sufficient decent condition holds.*

Proof. We use proof by induction from Eq.(3). We know that for $k = 0$ it is hold. Suppose that it is true for $k \geq 1$. Multiply Eq.(3) by g_k^T,

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

Divide by $\|g_k\|^2$ indicated that;

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

using Eq.(6), we have

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

This implies that,

$$-\sum_{j=0}^{k-1}\sigma^j \le \frac{g_k^T d_k}{\|g_k\|^2} \le -2+\sum_{j=0}^{k-1}\sigma^j,$$

since $\sum\limits_{j=0}^{k-1}\sigma^j \le \frac{1-\sigma^k}{1-\sigma}$ implies that,

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

Since $\sigma^k < 1$, when $\sigma \le \frac{1}{2}$ we have,

$$\frac{1-\sigma^k}{1-\sigma} < 2.$$

Thus we obtain $g_k^T d_k \le -c\|g_k\|^2$, where $c \in (0,1)$. The proof is completed. □

Theorem 3.4. *Suppose Assumption 1 is true, consider any CG method of form Eq.(2) and Eq.(3), where α_k satisfied SWP line search and the sufficient descent condition holds, then* $\lim\limits_{k\to\infty}\|g_k\| = 0$.

Proof. Suppose Eq.(15) holds true. Squaring both sides for Eq.(3) we obtained,

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

From (9) and (11),

$$\|d_k\|^2 \le \|g_k\|^2 - 2\frac{\|g_k\|^2}{\|d_{k-1}\|^2}g_k^T d_{k-1} + \left(\frac{\|g_k\|^2}{\|d_{k-1}\|^2}\right)^2\|d_{k-1}\|^2.$$

Since

$$-2\frac{\|g_k\|^2}{\|d_{k-1}\|^2}g_k^T d_{k-1} \le 2\frac{\|g_k\|^2}{\|d_{k-1}\|^2}|g_k^T d_{k-1}| \le 2\sigma(2-c)\frac{\|g_k\|^4}{\|d_{k-1}\|^2},$$

then

$$\|d_k\|^2 \le \|g_k\|^2 + 2\sigma(2-c)\frac{\|g_k\|^4}{\|d_{k-1}\|^2} + \frac{\|g_k\|^4}{\|d_{k-1}\|^2},$$

divide by $\|g_k\|^4$,

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \frac{1}{\|g_k\|^2} + \frac{2\sigma(2-c)}{\|d_{k-1}\|^2} + \frac{1}{\|d_{k-1}\|^2}.$$

Let $r = 2\sigma(2-c) + 1$,

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \frac{1}{\|g_k\|^2} + \frac{r}{\|d_{k-1}\|^2}.$$

Since $\frac{1}{\|g_0\|} = \frac{1}{\|d_0\|}$, we indicate that

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq (r+1)\sum_{i=0}^{\infty} \frac{1}{\|g_i\|^2},$$

using Eq.(15) we get, $\sum_{i=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} = \infty$, which contradicts Eq.(14). The proof is completed. □

4. Numerical results and discussions

To analyze the efficiency of the new method, We used here (Bongartz and Ingrid, 1995)
,(Andrei and Neculai, 2008)
, and (Adorio and Diliman, 2005)
test problems to analyze the efficiency of the new method. We performed a comparison with other CG methods, including VHS, $NPRP$, and $DPRP$. The tolerance ϵ is selected equal to 10^{-6} for all algorithms to investigate the rapidity of the iteration of these algorithm towards the optimal solution and the gradient value as the stopping criteria. Here, the stopping criteria considered $\|g_k\| \leq 10^{-6}$ for all algorithms. We used Matlab 7.9 subroutine program, with CPU processor Intel (R) Core (TM), i5 CPU, and 4GB DDR2 RAM . The performance results are shown in Figures 1 and 2 respectively, using a performance profile introduced by (Dolan et al, 2002).

In figures 1 and 2, we compare the new method with $NPRP$, WYL, and VHS methods based on CPU time and number of iteration to reach the optimum solution. It is clearly to see that the graph of AIM method above all of others and the best performance according to Dolan and More is the top right one, so the new method is an efficient method and better than other by using above testing functions.

Table 1: A list of problem functions which used under strong wolf condition $\delta = 0.01$ and $\sigma = 0.1$.

Function	Number of variables	Initial points		
Extended White and Holst	500,1000,5000	(-1.2,1,-1.2,...),	(5,5,...,5),	(10,...,10)
Extended Rosenbrock	500,1000,5000	(-1.2,1,-1.2,...),	(5,5,...,5),	(10,...,10)
Ext. Freudenstein and Roth	500,1000,5000	(0.5,...,.5),	(1,1,...,1),	(2,2,...,2)
Extended Beale	500,1000,5000	(1,1,...,1),	(5,5,...,5),	(10,...,10)
Perturbed Quadratic	500,1000,5000	(.5,.5,...,.5),	(1,1,...,1),	(2,2...,2)
Extended Himmelblau	500,1000,5000	(1,1,...,1),	(5,5,...,5),	(10,...,10)
Diagonal 2	500,1000,5000	(.2,.2,...,.2),	(.25,...,.25),	(.5,.5,...,5)
Nconscomp function	500,1000,5000	(1,1,...,1),	(-1,-1,...,-1),	(-2,-2...,-2)
Extended DENSCHNB	500,1000,5000	(1,1,...,1),	(5,5,...,5),	(10,...,10)
Shallow function	500,1000,5000	(-2,-2,...,-2),	(2,2,...,2),	(5,5,...,5)
Quadratic QF2	500,1000	(.5,.5,...,.5),	(2,2,...,2),	(5,5,...,5)
Ext. quadratic penalty QP2	100, 200	(2,2,...,2),	(5,5,...,5),	(10,...,10)
DIXMAANA function	500,1000,5000	(2,2,...,.2),	(5,5,...,5),	(10,...,10)
DIXMAANB function	500,1000,5000	(-2,-2,...,-2),	(-1,-1,...,-1),	(0,0,...,0)
NONDIA function	500,1000,5000	(-2,-2,...,-2),	(-1,-1,...,-1),	(0,0...,0)
Ext. Block Diagonal BD1	500,1000,5000	(1,1,...,1),	(-1,-1,...,-1),	(2,2,...,2)
DQDRTIC function	500,1000,5000	(-1,-1,...,-1),	(1,1,...,1),	(2,2,...,2)
Diagonal4 function	500,1000,5000	(1,1,...,1),	(5,5,...,5),	(10,...,10)
Raydan 2 function	500,1000,5000	(1,1,...,1),	(5,5,...,5),	(10,...,10)
Extended Cliff function	500,1000	(1,1,...,1),	(5,5,...,5),	(10,...,10)

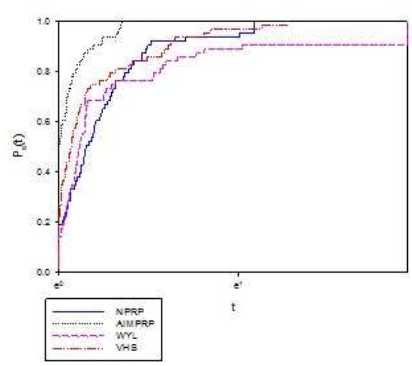

Figure 1: Performance profile based on the CPU time.

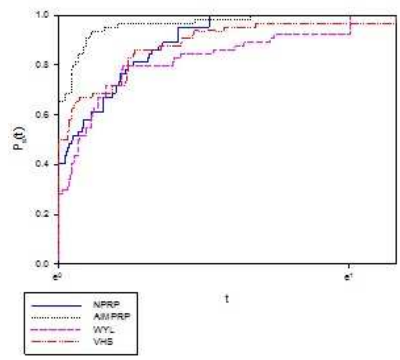

Figure 2: Performance profile based on the number of iteration.

5. Conclusion

In this paper, we presented a new modification of NPRP method, the new method has a global convergence under exact and strong Wolfe line searches, it passes the sufficient descent condition when $\sigma \leq \frac{1}{2}$. The numerical results show the new method is better than other modern CG methods. In future, we plane to proof the new method under several lines searches such weak Wolfe line search and Armijo line search.

Acknowledgments

The authors are grateful for all these supported and improve our paper, also we would like to thank the University of Malaysia Terengganu (UMT) for funding under FRGS Vot no. 59256.

References

Adorio, Ernesto, P. and Diliman, U. P. (2005). *Mvf-multivariate test functions library in c for unconstrained global optimization.*

Al-Baali, M. (1985). Descent property and global convergence of the fletcher-reeves method with inexact line search. *IMA Journal of Numerical Analysis*, (5.1):121–124.

Alhawarat, A., M. M. R. M. M. (2014). A new modification of nonlinear conjugate gradient coefficients with global convergence properties. *International Journal of Mathematical, Computational, Physical and Quantum Engineering*, (8(1)):54 – 60.

Andrei and Neculai (2008). An unconstrained optimization test functions collection. *Adv. Model. Optim*, (10.1):147–161.

Bongartz and Ingrid (1995). Cute: Constrained and unconstrained testing environment. *ACM Transactions on Mathematical Software (TOMS)*, (21.1):123–160.

Dai and Hong, Y. (2011). *Nonlinear conjugate gradient methods.* Wiley Encyclopedia of Operations Research and Management Science.

Fletcher, R. and Reeves, C. (1964). Function minimization by conjugate gradients. *Computer journal*, (7):149–154.

Hager, William, W. and Zhang, H. (2006). A survey of nonlinear conjugate gradient methods. *Pacific journal of Optimization*, (2.1):35–58.

Ibrahim, M. A. Hery, M. M. and Leong, W. (2014). *The Hybrid BFGS-CG Method in Solving Unconstrained Optimization Problems.* Hindawi Publishing Corporation, 2014 edition.

Mamat and Mustafa (2010). A new conjugate gradient coefficient for unconstrained optimization. *t. J. Contemp. Math. Sciences*, (5.29):1429–1437.

Polak, B. and Ribiere, G. (1969). Note surla convergence des methodes de directions conjuguees. *Rev. Fr. Aut-om. Inform. Rech. Oper.,3e Annee*, (16):35–43.

Rivaie, M., M. M. M. I. J. L. W. (2012). A new class of nonlinear conjugate gradient coefficients with global convergence properties. *Appl.Math. Comput*, (218):11323–11332.

Wei, Z., You, S., and Liu, L. (2006). The convergence properties of some new conjugate gradient methods. *Appl. Math. Comput.*, (183):1341–1350.

Zhang, L. (2009). An improved wei-yao-liu nonlinear conjugate gradient method for optimization computation. *Appl. Math. Comput.*, (215):2269–2274.

Zoutendijk, G. (1970). Nonlinear programming computational methods. *J. Abadie (Ed.), Integer and Nonlinear Programming, North-Holland, Amsterdam*, pages 37–86.