Performance comparison of a new hybrid conjugate gradient method under exact and inexact line searches

N H A Ghani¹, N S Mohamed¹, N Zull¹, S Shoid¹ M Rivaie² and M Mamat¹
¹Faculty of Informatics & Computing, Universiti Sultan Zainal Abidin (Kampus Gong Badak), 21300 Kuala Terengganu, Terengganu, Malaysia
²Department of Computer Science & Mathematics, Universiti Teknologi MARA (Kampus Kuala Terengganu), 21080 Kuala Terengganu, Terengganu, Malaysia

E-mail: hamizah.ghani@yahoo.com

Abstract. Conjugate gradient (CG) method is one of iterative techniques prominently used in solving unconstrained optimization problems due to its simplicity, low memory storage, and good convergence analysis. This paper presents a new hybrid conjugate gradient method, named NRM1 method. The method is analyzed under the exact and inexact line searches in given conditions. Theoretically, proofs show that the NRM1 method satisfies the sufficient descent condition with both line searches. The computational result indicates that NRM1 method is capable in solving the standard unconstrained optimization problems used. On the other hand, the NRM1 method performs better under inexact line search compared with exact line search.

1. Introduction

Conjugate gradient (CG) method is used to solve the following unconstrained optimization problem:

\[
\min_{x \in \mathbb{R}^n} f(x),
\]

where the objective function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is continuously differentiable. A sequence \( \{x_k\} \) is generated by an iterative method known as

\[
x_{k+1} = x_k + \alpha_k d_k,
\]

where \( \alpha_k \) and \( d_k \) is a step size and search direction respectively at the current iterative point \( x_k \in \mathbb{R}^n \). It is crucial to set the values of \( \alpha_k \) such that it can either be determined by exact or inexact line search.

The exact line search uses the following condition:

\[
f(x_{k+1}) = f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k)
\]

Meanwhile, inexact line search which particularly known as strong Wolfe-Powell line search uses the following conditions with \( 0 < \delta < \sigma < 1 \):

\[
f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k,
\]
while the search direction $d_k$ in (1.2) is defined as;

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

where $g_k$ is the gradient of $f(x)$ at point $x_k$ and $\beta_k$ refers to any CG parameter discussed in [1-6].

In this paper, a new hybrid CG method is introduced. Details explanation of the proposed method is discussed clearly in Section two. In Section three, the proofs of sufficient descent conditions for both line searches are shown. The computational results of the new method tested on a set of test problems are reported graphically in Section 4. To sum up all works, the conclusion of this paper is presented in the last section.

2. New hybrid conjugate gradient method

A major problem of the existing CG methods is that it can be very slow on certain types of unconstrained optimization problems. Therefore, there is a need to improve the efficiency of those methods in terms of the iteration numbers and central processing unit (CPU) per times. This section introduces a new hybrid CG method, namely NRM1 method. The hybridization process happened when two or more classical CG algorithms are combined. The combination of classical CG methods used in this paper are known as HS [7] and NRM [8] methods.

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

$$\beta_k^{NRM} = \begin{cases} \|g_k\|^2 - m_k^T g_k g_{k-1} \\ \frac{\|d_{k-1}\|^2}{m_k}, \text{ if } \|g_k\|^2 > m_k g_k^T g_{k-1} \\ 0, \text{ otherwise} \end{cases}$$

Many researchers discussed that HS method has a significant characteristic, where the conjugate relation of $d_k^T (g_k - g_{k-1}) = 0$ is always formed with any line precision used [9-12]. In addition, the HS method is accepted to be among the efficient methods because of its restart capabilities in order to avoid jamming and encounter bad direction [13-14]. However, the practical performance of HS method is classified as worst as the method did not converge for certain unconstrained optimization test functions. Thus, the development of NRM1 method will combine all the good criteria of HS and NRM methods, as well as having better numerical performance. Combining (2.1) and (2.2), the new NRM1 method is in the form of;

$$\beta_k^{NRM1} = \begin{cases} \beta_k^{HS}, \text{ if } \|g_k\|^2 > \|g_k^T g_{k-1}\| \\ \beta_k^{NRM}, \text{ otherwise} \end{cases}$$

Simplifying (2.3), the NRM1 method can be divided into two cases:

Case 1. If $\beta_k^{NRM1} = \beta_k^{HS}$, then under the condition $\|g_k\|^2 > \|g_k^T g_{k-1}\|$, NRM1 is written as

$$\beta_k^{NRM1} = \frac{\|g_k\|^2 - g_k^T g_{k-1}}{d_{k-1}^T (g_k - g_{k-1})} \leq \frac{\|g_k\|^2}{d_{k-1}^T g_k - d_{k-1}^T g_{k-1}}$$

Case 2. If $\beta_k^{NRM1} = \beta_k^{NRM}$, then by referring [8], NRM1 is simplified as follows:
\[ 0 \leq \beta_k^{NRM1} \leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2}. \]  

(2.5)

The algorithm of new hybrid CG method used in this study is given as follow:

**Step 1:** Set \( k = 0 \) and select an initial point \( x_0 \in \mathbb{R}^n \).

**Step 2:** Compute the new CG coefficient, \( \beta_k^{NRM1} \) based on (5).

**Step 3:** Compute the search direction, \( d_k = -g_k + \beta_k^{NRM1} d_{k-1} \). If \( \|g_k\| = 0 \), then stop.

**Step 4:** Compute \( \alpha_k \) by using exact and inexact line search.

**Step 5:** Update new point based on iterative formula (1.2).

**Step 6:** Convergent test and stopping criteria.

If \( f(x_{k+1}) < f(x_k) \) and \( \|g_{k+1}\| \leq \varepsilon \), then stop. Otherwise go to Step 1 with \( k = k + 1 \).

3. **Convergence analysis**

Every CG method must satisfy the descent property to ensure that it is convergent. The sufficient descent condition is defined as follows:

\[ g_k^T d_k \leq -c\|g_k\|^2 \text{ for } \forall k \geq 0 \text{ and } c > 0 . \]  

(3.1)

### 3.1. Sufficient descent condition of NRM1 method with exact line search

The following theorem is useful to show that the NRM1 method with exact line search satisfies (3.1).

**Theorem 1.** Consider NRM1 method with the search direction (1.6) is determined by the exact line search (1.3), then sufficient descent condition (3.1) holds true for all \( k \geq 0 \).

**Proof.** Theorem 1 is proved by induction. The condition (3.1) is true if \( k = 0 \) as \( g_0^T d_0 = -C\|g_0\|^2 \). Then, suppose that (3.1) is also true for \( k \geq 1 \). Set \( k = k + 1 \) and multiply (1.6) by \( g_{k+1}^T \), implies

\[ g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_{k+1}^{NRM1} g_{k+1}^T d_k \]  

(3.2)

It is known \( g_{k+1}^T d_k = 0 \) that using the exact line search. Thus, \( g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 \). Hence, the sufficient descent condition is holds. The proof is completed.

### 3.2. Sufficient descent condition of NRM1 method with inexact line search

The following discussion shows the proof of sufficient descent condition for NRM1 method with inexact line search, particularly on the strong Wolfe-Powell line search. **Case 1** is proved by Theorem 2. Meanwhile, **Case 2** is proved by Lemma 1 and Theorem 3.

**Theorem 2.** Consider NRM1 method with the search direction (1.6) is determined by (1.4) and (1.5), with \( \sigma < \frac{1}{8} \), then sufficient descent condition (3.1) holds true for all \( k \geq 0 \).

**Proof.** Theorem 2 is proved by induction. The condition (3.1) is true if \( k = 0 \) as \( g_0^T d_0 = -C\|g_0\|^2 \). Hence, suppose that (3.1) is also true for \( k \geq 1 \). **Case 1** can be divided into two categories, as follows:

**Case 1a.** If \( g_{k+1}^T d_k \leq 0 \), then (3.2) becomes

\[ g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_{k+1}^{NRM1} g_{k+1}^T d_k < 0 \]  

(3.3)
Case (1b). If \( g^T_{k+1}d_k > 0 \), then divides both sides of (3.2) by \( \|g_{k+1}\|^2 \) implies that,

\[
\frac{g^T_{k+1}d_{k+1}}{\|g_{k+1}\|^2} = -1 + \frac{\beta_{k+1}^\text{NRM} g^T_{k+1}d_k}{\|g_{k+1}\|^2}, \tag{3.4}
\]

Substitute (2.4) and apply (1.4) and (1.5) into (3.4),

\[
\frac{g^T_{k+1}d_{k+1}}{\|g_{k+1}\|^2} \leq -1 + \frac{\sigma g^T_{k}d_k}{(1-\sigma)\|g_{k}\|^2} = -1 + \frac{\sigma}{(1-\sigma)} < 0, \quad \text{when} \quad \sigma < \frac{1}{8} \tag{3.5}
\]

Let \( c = 1 - \frac{\sigma}{(1-\sigma)} \), thus (3.5) also can be written as \( g^T_{k+1}d_{k+1} \leq -c\|g_{k+1}\|^2 \). Hence, the sufficient descent condition holds true, and the proof is completed.

**Lemma 1.** Consider NRM1 method with the search direction (1.6) is determined by (1.4) and (1.5), with \( \sigma < \frac{1}{8} \), then for all \( k \geq 0 \), we have \( \frac{\|g_k\|}{\|d_k\|} \leq 2 \).

**Theorem 5.** Consider NRM1 method with the search direction (1.6) is determined by (1.4) and (1.5), with \( \sigma < \frac{1}{8} \), then sufficient descent condition (3.1) holds true for all \( k \geq 0 \).

**Proof.** Theorem 5 is proved by induction. The condition (3.1) is true if \( k = 0 \) as \( g^T_0d_0 = -C\|g_0\|^2 \).

Then, suppose that (3.1) is also true for \( k \geq 1 \). Dividing both sides of (3.2) by \( \|g_{k+1}\|^2 \) implies that,

\[
\frac{g^T_{k+1}d_{k+1}}{\|g_{k+1}\|^2} = -1 + \beta_{k+1}^\text{NRM} \frac{g^T_{k+1}d_k}{\|g_{k+1}\|^2} \tag{3.6}
\]

From the strong Wolfe-Powell condition and absolute value properties, (3.6) becomes;

\[
-1 + \sigma \beta_{k+1}^\text{NRM} \frac{\|g_k\|^2}{\|g_{k+1}\|^2} \leq \frac{g^T_{k+1}d_{k+1}}{\|g_{k+1}\|^2} \leq -1 - \sigma \beta_{k+1}^\text{NRM} \frac{\|g_k\|^2}{\|g_{k+1}\|^2} \tag{3.7}
\]

For Case 2, substitute (2.5) and apply Lemma 2 in (3.7) becomes

\[
-1 + 4\sigma \frac{g^T_{k}d_k}{\|g_{k}\|^2} \leq \frac{g^T_{k+1}d_{k+1}}{\|g_{k+1}\|^2} \leq -1 - 4\sigma \frac{g^T_{k}d_k}{\|g_{k}\|^2} \tag{3.8}
\]

Repeating this process and using the fact \( g^T_0d_0 = -\|d_0\|^2 \), (3.45) can be written as

\[
- \sum_{i=0}^{k} (4\sigma)^i \leq \frac{g^T_{k+1}d_{k+1}}{\|g_{k+1}\|^2} \leq -2 + \sum_{i=0}^{k} (4\sigma)^i \tag{3.9}
\]

Since \( \sum_{i=0}^{k} (4\sigma)^i \leq \frac{1-(4\sigma)^k}{1-4\sigma} \) and if \( c = 2 + \frac{1-(4\sigma)^k}{1-4\sigma} \), then \( c - 2 \leq \frac{g^T_{k+1}d_{k+1}}{\|g_{k+1}\|^2} \leq -c \).
Hence, it shows that \( g_k^T d_{k+1} \leq -c \| g_{k+1} \|^2 \), where \( c \in (0,1) \). This proved that the result holds true for \( k \geq 1 \). Therefore, the sufficient descent condition (3.1) holds. The proof is completed.

4. Numerical results and discussion

This section presents the numerical result of the NRM1 method in order to test the efficiency of the new method in solving unconstrained optimization problems. The NRM1 method is evaluated under the exact and inexact line searches. Then, the performances of NRM1 method within both line searches are compared. The results are analyzed based on the number of iterations and CPU times. The tests are implemented on a set of fifteen test problems taken from [15], described in Table 1. As suggested by Hillstrom [16], the condition \( \| g_k \| \leq 10^{-6} \) is used as the stopping criterion. The calculation is also terminated and considered fail to approach the solution point if the iterations exceed ten thousands. All calculations are done by using Matlab R2011b on a PC with Intel(R), Core(TM), i7-4712MQ 2.30GHz and 8GB RAM memory.

| No. | Problems          | Dimension/s | Initial Points |
|-----|-------------------|-------------|----------------|
| 1   | Arwhead           | 2, 4, 10    | (4,4,…,4), (14,14,…,14), (24,24,…,24), (40,40,…,40) |
| 2   | Booth             | 2           | (1,1), (5,5), (10,10), (20,20) |
| 3   | Diagonal 4        | 2, 500, 1000, 5000, 10000 | (3,3,…,3), (10,10,…,10), (50,50,…,50), (80,80,…,80) |
| 4   | DIXMAANA          | 3, 900, 3000, 6000, 9000 | (2,2,…,2), (10,10,…,10), (15,15,…,15), (35,35,…,35) |
| 5   | DIXMAANB         | 3, 900, 3000, 6000, 9000 | (-5,-5,…,-5), (-2,-2,…,-2), (2,2,…,2), (5,5,…,5) |
| 6   | Extended Beale    | 2, 500, 1000, 5000, 10000 | (-1,-1,…,-1), (0.5,0.5,…,0.5), (1,1,…,1), (2,2,…,2) |
| 7   | Extended Cliff    | 2, 500, 1000, 5000, 10000 | (10,10,…,10), (20,20,…,20), (50,50,…,50), (80,80,…,80) |
| 8   | Extended Himmelblau | 2, 500, 1000, 5000, 10000 | (1,1,…,1), (5,5,…,5), (20,20,…,20), (25,25,…,25) |
| 9   | Extended Rosenbrock | 2, 500, 1000, 5000, 10000 | (2,2,…,2), (5,5,…,5), (10,10,…,10), (20,20,…,20) |
| 10  | Hager             | 2, 4, 10, 100 | (2,2,…,2), (5,5,…,5), (10,10,…,10), (20,20,…,20) |
| 11  | NONDIA            | 2, 500, 1000, 5000, 10000 | (5,5,…,5), (15,15,…,15), (40,40,…,60), (60,60,…,60) |
| 12  | Raydan 1          | 2, 4, 10, 100 | (2,2,…,2), (5,5,…,5), (10,10,…,10), (20,20,…,20) |
| 13  | Six Hump Camel    | 2           | (1,1), (5,5), (10,10), (15,15) |
| 14  | Shallow           | 2, 10, 100, 500, 1000 | (12,12,…,12), (22,22,…,22), (32,32,…,32), (62,62,…,62) |
| 15  | Zettl             | 2           | (1,1), (10,10), (15,15), (25,25) |

Performance comparisons of NRM1 method under both line searches are analysed by using the performance profile suggested by Dolan and Morè [17]. Figure 1 and Figure 2 show the performance profile of NRM1 method, corresponding to the number of iterations and CPU times respectively. The right side of graph illustrates the percentage of the test problems that are successfully solved by the NRM1 method. Meanwhile, the left side gives the percentage of the test problems with the fastest method. Interestingly, there is a slightly large comparison between both line searches. The performance results by using inexact line search is more effective compare to the exact line search, both in terms of iteration numbers and CPU times. This is because, it can solves a large number of unconstrained optimization problems and converge faster to get the solutions.
Figure 1. Performance Profile with respect to the number of iteration.

Figure 2. Performance Profile with respect to the CPU times.

5. Conclusion
In this paper, a new hybrid CG method known as NRM1 is introduced. While maintaining its convergence properties, result show that NRM1 method is superior under both line searches and particularly has a better performance under inexact line search. For further research, the new CG method will contribute to the knowledge in the sense of its application.

Acknowledgments
Authors would like to thank the editors and referees for their suggestion and comments. The authors are also thankful to the Ministry of Higher Education of Malaysia for the funding of this research via MyPhD KPM.

References
[1] ‘Aini N, Rivaie M and Matam M 2016 Aip Conf. Proc. 1787 080019
[2] Mohamed N S, Mamad M, Mohamad F S and Rivaie M 2016 Jurnal Teknologi Science & Engineering, 78 131-136
[3] Rivaie M, Mamad M, Ismail M and Fauzi M 2011 Aus. J. Bas. Appl. Sci. 5, 947–951
[4] Rivaie, M, Mamad M and Abashar A 2015 Appl. Math. and Comput. 268, 1152–1163
[5] Shoid S, Rivaie M and Mamad M 2016 Aip Conf. Proc. 1739
[6] Zull N, Rivaie M, Mamad M, Salleh Z and Amani Z 2015 Applied Math. Science 9 3105-3117.
[7] Hestenes M R and Steifel E 1952 J. Res. Nat. Bur. Stand. 49, 409-436.
[8] Ghani N H A, Rivaie M and Mamad M 2016 Aip Conf. Proc. 1739 020076
[9] Salleh Z and Alhawarat A 2016 Journal of Inequalities and Applications 110(1)
[10] Xu X and Kong F Y 2016 SpringerPlus, 5(1), 881
[11] Hager W W and Zhang H 2006 Pacific journal of Optimization 2(1), 35-58
[12] Zhang L 2009 Computational & Applied Mathematics, 28(1), 111-133
[13] Abdullahi I and Ahmad R 2016. Indian Journal of Science and Technology, 9(41)
[14] Andrei N 2013 Studies in Informatics and Control 16(4) 333-352
[15] Andrei N 2008 Advanced Modelling and Optimization 10, 147-161
[16] Hilstrom K E 1977 A.C.M. Trans. Maths. Softw. 3, 4, 305-315
[17] Dolan E and More J J 2002 Math. Programming 91, 201-213