A New Restating Criterion for FR-CG Method with Exact and Inexact Line Searches

Maha S. Younis
College of Education
University of Mosul, Iraq

Received on: 11/09/2007  Accepted on: 04/03/2008

ABSTRACT
A new restarting criterion for FR-CG method is derived and investigated in this paper. This criterion is globally convergent whenever the line search fulfills the Wolfe conditions. Our numerical tests and comparisons with the standard FR-CG method for large-scale unconstrained optimization are given, showing significantly improvements.

Keywords: Unconstrained optimization, FR-CG method, restarting criterion, line search, Wolfe conditions.

1. INTRODUCTION:
The classical conjugate gradient method to minimize a non linear function \( f(x) \) of the vector variable \( x= (x_1, x_2, \ldots, x_n)^T \) is an iterative method defined by

\[
\begin{align*}
x_{i+1} &= x_i + \alpha_i d_i \quad \ldots(1) \\
d_1 &= -g_i \quad \ldots(2) \\
\text{and} \\
d_{i+1} &= -g_{i+1} + \beta_i d_i \quad \ldots(3)
\end{align*}
\]

where \( g_i = \nabla f(x_i) \), \( \alpha_i \) is a line search parameter, and
with $\gamma_i = g_{i+1} - g_i$ the method was originally proposed by Hestenes and Stiefel [Hestenes and Stiefel, 1952] to solve a systems of linear equations, and first applied to nonlinear optimization problems by Fletcher and Reeves [Fletcher and Reeves, 1964].

In the original Fletcher-Reeves paper, the parameter $\beta_i$ defined by (4) is redefined by:

$$\beta^{FR}_i = \frac{g_{i+1}^T g_i}{g_i^T g_i}$$

(5)

The definitions (4) and (5) are identical if $\alpha_i$ is chosen to minimize $f(x)$ along $d_i$ and $f(x)$ is quadratic.

Polak and Ribiere [Polak and Ribiere, 1969] suggested a $\beta_i$ defined by:

$$\beta^{PR}_i = \frac{g_{i+1}^T y_{i+1}}{g_i^T g_i}$$

(6)

which is identical to (4) whenever $\alpha_i$ is chosen to minimize $f(x)$ along $d_i$, independent of any assumption.

Shanno [Shanno, 1978] noted that the search direction (3) was equivalent to:

$$d_{i+1} = \left( I - \frac{\delta_i y_i^T + y_i \delta_i^T}{\delta_i^T y_i} \right) + \left( I + \frac{y_i^T y_i}{\delta_i^T y_i} \right) g_{i+1}$$

(7)

$$\delta_i = \alpha_i d_i$$

whenever $d_i^T g_{i+1} = 0$. The last condition is simply the condition that $\alpha_i$ minimize $f(x)$ along $d_i$, an advantage of (7) over (3) is that under much looser line search criteria than exact line minimization, the direction is a descent direction, while all the above algorithms reduce to the same algorithm under the assumption of exact line minimization and a quadratic $f(x)$. A complicated algorithm based on (7), using self scaling, Beale restarts [Beale, 1972] and Powell’s restart criterion [Powell, 1977] has been implemented [Shanno and Phua, 1980], and shown to be generally numerically far more efficient than any of the standard algorithms using (3) with various choices of $\beta_i$.

Further, the algorithm has been shown to converge to a stationary point of $f(x)$ [Shanno, 1978] under loose line search criteria for convex functions, but has not been shown convergent for general functions satisfying the conditions that:

$F(x)$ has continuous second partial derivatives ... (8)
And the set $x$ defined by:

\[ \{ x \mid f(x) < f(x_1) \} \] is bounded \hspace{1cm} \text{(9)}

Zoutendijk (1970) showed convergence of the Fletcher-Reeves conjugate gradient method, corresponding to the choice of $\beta_1$, defined by (5), for such functions which have also recently been shown by Powell (1983).

Powell's paper, however, also shows that for $\beta_1$ chosen to satisfy (4) rather than (5), even with exact line searches, there exist functions satisfying (8) and (9) where the sequence (1)-(3) cycles infinitely.

Furthermore, on the sequence of points for which cycling occurs, $g(x)$ is bounded away from zero.

It is the purpose of this note to show that convergence proof for the Fletcher-Reeves method may be used to guarantee convergence to stationary point for any conjugate gradient method. Numerical results testing the proposed modification on the algorithm of Shanno and Phua show that the efficiency of the modified algorithm is no worse than the original algorithm, and is sometimes better.

Further, test results indicate potential real improvement of the original algorithm may be achieved for at least some large problems. As large problems are the problems for which conjugate gradient methods have been devised, the test appears to have computational as well as theoretical utility [Shanno, 1985].

The work of Hestenes and Stiefel, (1952) presents a choice for $\beta_1$ closely related to the Polak and Ribiere scheme:

\[ \beta_1^{\text{HS}} = y_1^T g_1 / y_1^T d_1 \] \hspace{1cm} \text{(10)}

If $\alpha_i$ is obtained by an exact line search, then by (3) we have:

\[ y_i^T d_i = (g_{i+1} - g_i)^T d_i = g_i^T d_i - e_i^T g_i \] \hspace{1cm} \text{(11)}

Hence $\beta_1^{\text{HS}} = \beta_1^{\text{PR}}$ when $\alpha_i$ is obtained by an exact line search.

More recent nonlinear conjugate gradient algorithms include the conjugate descent algorithm of Fletcher (1987) the scheme of Liu and Storey [1991], and the scheme of Dai and Yuan, (1999), (See also the survey article of Hager and Zhang, (2006). The scheme of Dai and Yuan corresponds to the following choice for the update parameter [Hager and Zhang, 2006]. By:

\[ \frac{\| g_{i+1} \|^2}{d_i^T y_i} \]

2. Restarting Criteria for a CG-Algorithm:
In the implementation of many CG-algorithms, one may often meet the difficulty that the search direction of some iteration is very poor. For example, the Newton direction is not well-defined if the Hessian of the objective function is singular but not positive, the Newton's direction is not necessarily a descent direction. Also PR-CG is now believed to be one of the most efficient CG-methods even for strictly convex quadratic function. However, PR-CG method with strong Wolfe condition may produce an uphill search direction is poor, a simple way is to restart. The method with $g_k$ is to guarantee the global convergence of the method. In this section, we can investigate and derive a new restarting criterion restart FR-CG and still obtain the global convergence property.

CG-methods are usually implemented with restarts after $n$ iterations, to match the quadratic model and in order to avoid the effects of an accumulation of errors. It was shown by Cohen (1972) that several restarted CG-methods have $n$-step quadratic convergence. It was established by Crounder and Wolfe (1972) that if restating is not employed for general functions, the convergence of CG-methods will only be linear: they also came to the conclusion that convergence is not better than linear for quadratic functions. Again Powell (1976) showed that for a convex quadratic function the convergence rate is linear. Fletcher and Reeves (1964) suggested restarting their algorithm every $n$ iterations where $n$ is the number of variables. Their standard reset was:

$$d_i = -g_i$$ for $i = 1, n, 2n, \ldots$ \hspace{1cm} (13)

The following remarks show that the Fletcher-Reeves algorithm may be inefficient for several iterations if a search direction $d_i$ occurs that is almost orthogonal to the steepest decent direction $-g_i$. We let $\theta_i$ be the angle between $d_i$ and $-g_i$, the definition:

$$d_i = -g_i + \beta_i d_{i-1}$$ \hspace{1cm} (14)

and the orthogonality of $g_i$ to $d_{i-1}$. This is useful because it gives the equation:

$$||d_i|| = \sec \theta_i ||g_i||$$ \hspace{1cm} (15)

Further, if $i$ is replaced by $(i + 1)$ in the figure, we find the identity:

$$\beta_{i+1} ||d_i|| = \tan \theta_{i+1} ||g_{i+1}||$$ \hspace{1cm} (16)
A New Restarting Criterion for FR-CG Method with…

Fig. (1)...The definition of \( \Theta_i \)

We may eliminate \( \| d_i \| \) from equations (15) and (16) and substitute the definition \( \beta_i = \frac{\| g_i \|^2}{\| g_i-1 \|^2} \) to deduce the inequality:

\[
\tan \Theta_{i+1} = \sec \Theta_i \frac{\| g_{i+1} \|}{\| g_i \|} \\
\Rightarrow \tan \Theta_{i+1} = \sec \Theta_i \frac{\| g_{i+1} \|}{\| g_i \|} 
\]

...(17)

Now if \( \Theta_i \) is close to \( \pi/2 \), the iteration may take a very small step in which case the change \( (g_{i+1} - g_i) \) is small also. Thus the ratio \( \frac{\| g_{i+1} \|}{\| g_i \|} \) is close to one. It follows from inequality (17) that \( \Theta_{i+1} \) is close to \( \pi/2 \), so slow progress may occur again on the next iteration.

Numerical calculations, show that this inefficient behavior can continue for several iterations when \( \beta_i \) is defined by equation

\[
\beta_i^{FR} = \frac{\| g_i \|^2}{\| g_{i-1} \|^2} 
\]
demonstration of this effect.

Suppose that the early iterations of the algorithm have made \( \Theta_i \) positive, but that a region in the space of the variables has been reached where \( f(x) \) is the quadratic function:

\[
f(x) = x_1^2 + x_2^2 
\]

In this case the line search along \( d_i \) makes the ratio \( \frac{\| g_{i+1} \|}{\| g_i \|} \) equal to \( \sin \Theta_i \)

Therefore the first line of expression (17) shows that \( \Theta_{i+1} \) is equal to \( \Theta_i \). Thus the angle between the search direction and the steepest descent direction remains constant for all iterations, which is highly inefficient if \( \Theta_i \) is close to \( \pi/2 \). Note that this inefficient behavior is corrected by a steepest descent restart.

Alternatively, if expression \( \beta_i = \frac{\| g_i \|^2}{\| g_{i-1} \|^2} \) is used to define \( \beta_i \) then the iterations of the conjugate gradient method have never seemed to be less efficient than those of the steepest descent method. We used equations (15) and (16) to show that the behavior described in the last two paragraphs does not occur.

Now the definition of \( \beta_i \) provides the bound:

\[
\beta_{i+1} < \frac{\| g_{i+1} \|}{\| g_{i+1} - g_i \|} / \| g_i \|^2 
\]

...(19)

So the elimination of \( \| d_i \| \) from the two equations gives the inequality:

\[
\tan \Theta_{i+1} < \sec \Theta_i \frac{\| g_{i+1} - g_i \|}{\| g_i \|} 
\]

...(20)

99
It follows that, if $\theta_i$ is close to $1/2 \pi$ and if this causes the step from $x_i$ to $X_{i+1}$ to be so small that the change $(g_{i+1} - g_i)$ is much less than $\|g_i\|$, then the $\tan \theta_{i+1}$ is much less than $\sec \theta_i$.

Thus the search direction $d_{i+1}$ is turned towards the steepest descent direction. Inequality (20) is sufficiently powerful to prove the following convergence theorem which, in contrast to a similar theorem given by Polak (1971) does not require $f(x)$ to satisfy any convexity conditions.

2.1 A new restarting criterion for FR-CG method

In this section we are going to introduce a new descent condition to FR-CG method as:

**Theorem 2.1:** If $d_i^T y_i \neq 0$ and $d_{i+1} = -g_{i+1} + \tau d_i$, \hspace{1cm} \ldots (21)
\hspace{1cm} $d_0 = -g_0$ for any $\tau \in [\beta_{FR}^i, \max\{\beta_{FR}^i, 0\}]$,
\hspace{1cm} \ldots (22-a)

\begin{align*}
&\text{then } \frac{g_{i+1}^T d_i}{\|g_i\|^2} \leq \frac{1}{8} \left( \frac{\|g_i\|^2}{\|g_i\|^2} + \frac{2(g_{i+1}^T d_i)^2}{\|g_i\|^2} \right) \hspace{1cm} \ldots (22-b)
\end{align*}

**Proof:**
\begin{align*}
&d_{i+1} = -g_{i+1} + \beta_i d_i,
&g_{i+1}^T d_{i+1} = -\|g_{i+1}\|^2 + \frac{g_{i+1}^T d_i}{\|g_i\|^2} \|g_{i+1}\|^2 \hspace{1cm} \ldots (23)
\end{align*}

Where $\beta_i$ of Fletcher-Reeve:
\begin{align*}
&= - \frac{\|g_{i+1}\|^2 \|g_i\|^2 + g_{i+1}^T d_i \|g_{i+1}\|^2}{\|g_i\|^2}
\end{align*}

Let $u = 1/2 \|g_{i+1}\| \|g_i\|$ and $v = 2 g_{i+1}^T d_i \|g_{i+1}\|

We apply the inequality:
\begin{align*}
u^T v &\leq 1/2 (||u||^2 + ||v||^2) \\
\frac{1}{2} \|g_{i+1}\| \|g_i\| 2g_{i+1}^T d_i \|g_{i+1}\| &\leq \\
\frac{1}{2} \left[ 1/4 \|g_{i+1}\|^2 \|g_i\|^2 + 4(g_{i+1}^T d_i)^2 \|g_{i+1}\|^2 \right] \\
\therefore \frac{g_{i+1}^T d_i \|g_{i+1}\|^2}{\|g_i\|^2} &\leq 1/8 \|g_{i+1}\|^2 \|g_i\|^2 + 2(g_{i+1}^T d_i)^2 \|g_{i+1}\|^2 \\
\therefore \frac{g_{i+1}^T d_i}{\|g_i\|^2} &\leq 1/8 \|g_i\|^2 + 2(g_{i+1}^T d_i)^2
\end{align*}
Hence
\[ g_{i+1}^T d_i \leq 1/8\|g_i\|^2 + \frac{2(g_{i+1}^T d_i)^2}{\|g_i\|^2} \] ...(24)
Substitute (24) in the eq. (23) we get
\[ g_i^T d_{i+1} \leq -\|g_{i+1}\|^2 + \frac{g_{i+1}^T}{\|g_i\|^2} \left( \frac{1}{8\|g_i\|^2} + \frac{2(g_{i+1}^T d_i)^2}{\|g_i\|^2} \right) \] ...(25)
In the ELS \[ g_{i+1}^T d_i = 0 \] this implies that
\[ g_i^T d_{i+1} \leq -\|g_{i+1}\|^2 + 1/8\|g_{i+1}\|^2 \]
Hence we get eq. (22-a) but in the ILS the restart is represented by the eq. (22-b).

3. Numerical Results:

The numerical performance of the CG-methods is greatly improved by using restarts. The disadvantages of restarting according to (13) is that the immediate reduction in the objective function is usually less than that what it would be without restarts. Moreover it is inefficient of errors and has already affected the conjugacy property.

A restart direction different from (13) was proposed by Beale, (1972), which can be used to derive a sophisticated restart procedure. The merit of Beale's restarting direction is that it allows an increase in the immediate reduction of the function value when using CG-method to minimize a non quadratic function.

Powell (1977), also developed a new procedure for restarting CG-methods. He suggested a restart criterion whenever:
\[ \left| g_i^T g_{i+1}\right| \leq 0.2 \left| g_i^T g_{i+1}\right| \] ...(26)

The rationale behind this check is that successive gradients will be close to orthogonality. He also checked that the new search direction \( d_{i+1} \) will be sufficiently downhill, using the formula:
\[ d_{i+1}^T g_{i+1} \leq -0.8(g_{i+1}^T g_{i+1}) \] ...(27)

or again a restart will be initiated. Numerical experiments performed by Powell justified the parameter values of 0.2 and -0.8 quoted in (26) and (27).

However, Boland, et al. (1979) used Powell’s restarting criterion, (26) or (27) to restart his polynomial model:
\[ f[q(x)] = \frac{\gamma_1 q(x)+1}{\gamma_2 q(x)}, \quad \gamma_2 < 0 \] ...(28)
obtained by a special nonlinear scaling of a quadratic function has been considered by Tassopoulos and Story (1984), with an arbitrary search direction other than the steepest descent with evident success (Al-Bayati, 1993).

And we define some symbols we use in the tables:

NOI = The number of iterations.
NOF = The number of function evaluations.
ELS = Exact Line Searches.
ILS = Inexact Line Searches.

Finally, from our numerical results: Table (3.1) indicates that there are no improvement for the new proposed algorithm (for both exact and inexact line searches) either for NOI or NOF because the dimensions for these test functions are small (N= 4).

From Table (3.2) we have the percentage performance of the new proposed technique against 100% F/R for (100 ≤ N ≤ 500)

| F/R | ENR | INR |
|-----|-----|-----|
| NOI | NOF | NOI | NOF |
| 7621 | 37083 | 1799 | 5206 |
| 100% | 100% | 23.5% | 14% |

Also, from Table (3.3) we have: the percentage performance of the new proposed technique against 100% F/R for (600 ≤ N ≤ 1000)

| F/R | ENR | INR |
|-----|-----|-----|
| NOI | NOF | NOI | NOF |
| 7621 | 37083 | 1946 | 5839 |
| 100% | 100% | 21.5% | 15.7% |

Table (3.1): Comparison for FR-CG method with standard a new restarting criteria for (N = 4) only

| Fun.  | N | F/R | ENR | INR |
|-------|---|-----|-----|-----|
| Wood  | 4 | 40  | 62  | 47  |
| Wolfe | 4 | 11  | 13  | 13  |
| Non-Dia. | 4 | 19  | 20  | 23  |
| Edger | 4 | 6   | 11  | 4   |
| Rosen | 4 | 27  | 54  | 38  |
| Recip | 4 | 6   | 7   | 7   |
| Powell | 4 | 18  | 42  | 19  |
| Sum   | 4 | 6   | 6   | 4   |
| Cubic | 4 | 14  | 13  | 25  |
| Helical | 4 | 36  | 41  | 79  |
| Total | 183 | 562 | 269 | 259 |
Table (3.2): Comparison for FR-CG method with standard a new restarting criteria for \((100 \leq N \leq 500)\)

| Fun.  | N  | F/R  | ENR  | INR  |
|-------|----|------|------|------|
|       |    | NOI  | NOF  | NOI  | NOF  | NOI  | NOF  |
| Wood  |    |      |      |      |      |      |      |
|       | 100| 258  | 962  | 68   | 162  | 47   | 107  |
|       | 200| 405  | 1662 | 76   | 178  | 47   | 107  |
|       | 300| 974  | 4959 | 78   | 182  | 47   | 107  |
|       | 400| 1492 | 8429 | 80   | 186  | 47   | 107  |
|       | 500| 1153 | 6064 | 81   | 188  | 47   | 107  |
| Wolfe |    |      |      |      |      |      |      |
|       | 100| 49   | 101  | 38   | 79   | 47   | 95   |
|       | 200| 52   | 107  | 40   | 83   | 44   | 89   |
|       | 300| 56   | 116  | 38   | 82   | 44   | 89   |
|       | 400| 60   | 126  | 39   | 84   | 43   | 87   |
|       | 500| 65   | 137  | 42   | 91   | 43   | 87   |
| Non-Dia. | |      |      |      |      |      |      |
|       | 100| 30   | 100  | 26   | 98   | 39   | 114  |
|       | 200| 31   | 101  | 27   | 98   | 37   | 108  |
|       | 300| 19   | 78   | 31   | 106  | 33   | 99   |
|       | 400| 38   | 114  | 33   | 107  | 32   | 97   |
|       | 500| 36   | 111  | 33   | 107  | 46   | 121  |
| Edger |    |      |      |      |      |      |      |
|       | 100| 13   | 33   | 13   | 33   | 5    | 12   |
|       | 200| 13   | 33   | 13   | 33   | 5    | 12   |
|       | 300| 13   | 33   | 13   | 33   | 5    | 12   |
|       | 400| 13   | 33   | 13   | 33   | 5    | 12   |
|       | 500| 13   | 33   | 13   | 33   | 5    | 12   |
| Rosen |    |      |      |      |      |      |      |
|       | 100| 77   | 287  | 60   | 203  | 38   | 94   |
|       | 200| 77   | 287  | 60   | 203  | 38   | 94   |
|       | 300| 78   | 289  | 60   | 203  | 38   | 94   |
|       | 400| 78   | 289  | 60   | 203  | 38   | 94   |
|       | 500| 82   | 297  | 60   | 203  | 38   | 94   |
| Recip |    |      |      |      |      |      |      |
|       | 100| 7    | 20   | 7    | 20   | 7    | 22   |
|       | 200| 7    | 20   | 7    | 20   | 7    | 22   |
|       | 300| 8    | 22   | 8    | 22   | 7    | 22   |
|       | 400| 8    | 22   | 8    | 22   | 7    | 22   |
|       | 500| 8    | 22   | 8    | 22   | 7    | 22   |
| Powell|    |      |      |      |      |      |      |
|       | 100| 48   | 97   | 48   | 97   | 19   | 41   |
|       | 200| 49   | 99   | 49   | 99   | 19   | 41   |
|       | 300| 52   | 105  | 52   | 105  | 19   | 41   |
|       | 400| 52   | 105  | 52   | 105  | 20   | 43   |
|       | 500| 52   | 105  | 52   | 105  | 21   | 45   |
| Fun.   | N  | F/R NOI | NOF | ENR NOI | NOF | INR NOI | NOF |
|--------|----|---------|-----|---------|-----|---------|-----|
| Sum    | 100| 17   | 155 | 15      | 125 | 13      | 82  |
|        | 200| 19   | 149 | 20      | 145 | 20      | 133 |
|        | 300| 21   | 174 | 23      | 173 | 21      | 136 |
|        | 400| 23   | 199 | 27      | 206 | 18      | 123 |
|        | 500| 27   | 238 | 28      | 214 | 19      | 135 |
| Cubic  | 100| 14   | 45  | 14      | 45  | 25      | 62  |
|        | 200| 14   | 45  | 14      | 45  | 25      | 62  |
|        | 300| 14   | 45  | 14      | 45  | 25      | 62  |
|        | 400| 14   | 45  | 14      | 45  | 25      | 62  |
|        | 500| 14   | 45  | 14      | 45  | 25      | 62  |
| Helical| 100| 105  | 211 | 46      | 98  | 80      | 163 |
|        | 200| 200  | 401 | 46      | 98  | 80      | 163 |
|        | 300| 202  | 405 | 46      | 98  | 80      | 163 |
|        | 400| 205  | 411 | 46      | 98  | 80      | 163 |
|        | 500| 206  | 413 | 46      | 98  | 80      | 163 |
| Total  | 6561| 28379|    | 1799    | 5206| 1607    | 4106|
Table (3.3) : Comparison for FR-CG method with standard a new restarting criteria for (600 ≤ N ≤ 1000)

| Fun.  | N   | F/R NOI | NOF | ENR NOI | NOF | INR NOI | NOF |
|-------|-----|---------|-----|---------|-----|---------|-----|
| Wood  | 600 | 718     | 4189| 82      | 190 | 47      | 107 |
|       | 700 | 904     | 5171| 82      | 190 | 47      | 107 |
|       | 800 | 1164    | 6305| 82      | 190 | 47      | 107 |
|       | 900 | 1065    | 6695| 82      | 190 | 47      | 107 |
|       | 1000| 1035    | 7156| 83      | 192 | 47      | 107 |
| Wolfe | 600 | 70      | 146 | 43      | 91  | 43      | 87  |
|       | 700 | 76      | 160 | 43      | 94  | 42      | 85  |
|       | 800 | 83      | 180 | 44      | 98  | 42      | 85  |
|       | 900 | 91      | 199 | 46      | 103 | 42      | 85  |
|       | 1000| 100     | 219 | 48      | 105 | 42      | 85  |
| Non-Dia. | 600 | 30      | 100 | 33      | 107 | 46      | 121 |
|       | 700 | 37      | 112 | 33      | 107 | 45      | 119 |
|       | 800 | 48      | 135 | 33      | 107 | 45      | 119 |
|       | 900 | 61      | 160 | 33      | 107 | 40      | 110 |
|       | 1000| 76      | 190 | 33      | 107 | 43      | 115 |
| Edger | 600 | 14      | 35  | 14      | 35  | 5       | 12  |
|       | 700 | 14      | 35  | 14      | 35  | 5       | 12  |
|       | 800 | 14      | 35  | 14      | 35  | 5       | 12  |
|       | 900 | 14      | 35  | 14      | 35  | 5       | 12  |
|       | 1000| 14     | 35   | 14      | 35  | 5       | 12  |
| Rosen | 600 | 82      | 297 | 60      | 203 | 38      | 94  |
|       | 700 | 82      | 297 | 60      | 203 | 38      | 94  |
|       | 800 | 82      | 297 | 60      | 203 | 38      | 94  |
|       | 900 | 82      | 297 | 60      | 203 | 39      | 96  |
|       | 1000| 82     | 297 | 60      | 203 | 39      | 96  |
| Recip | 600 | 8       | 22  | 8       | 22  | 7       | 22  |
|       | 700 | 8       | 22  | 8       | 22  | 7       | 22  |
|       | 800 | 9       | 26  | 9       | 26  | 7       | 22  |
|       | 900 | 9       | 26  | 9       | 26  | 7       | 22  |
|       | 1000| 9      | 26    | 9       | 26  | 7       | 22  |
| Powell | 600 | 52      | 105 | 52      | 105 | 21      | 45  |
|       | 700 | 53      | 107 | 53      | 107 | 21      | 45  |
|       | 800 | 53      | 107 | 53      | 107 | 21      | 45  |
|       | 900 | 53      | 107 | 53      | 107 | 22      | 47  |
|       | 1000| 53    | 107   | 53      | 107 | 22      | 47  |
## 4. Conclusions:

According to our numerical results we have concluded that using the new restarting criteria (eqs. (22-a)and(22-b)) from both exact (ELS) and inexact line searches (ILS) instate of the standard restarting criterion (K=N) for F/R-CG method are very useful technique only for medium and large dimensionality test functions namely there are (75-85)% NOI improvement and (75-80)% NOF improvement for medium and large test functions.
REFERENCES

[1] Al-Bayati, A. Y. "A new family of self-scaling variable metric algorithms for unconstrained optimization", J. Educ. & Sci. 12, (1991).

[2] Al-Bayati, A. Y. "A New Nonquadratic Model for Unconstrained Nonlinear Optimization", Mu'tah Journal for Research and Studies, 8, (1995), pp.131-155.

[3] Beale, E. M. L. "A derivation of conjugate gradients", in : F. A. Lootsma, ed.. Numerical methods for nonlinear optimization Academic Press, London, (1972), 39-43.

[4] Boland, W. R. and Kowalik, J. S. "Extended conjugate gradient methods with restarts". Journal of Optimization Theory and Applications, 28, (1979), 1-9.

[5] Cohen, A. I. "Rate of Convergence of Several conjugate gradient algorithms", SIAM Journal of Numerical Analysis, 9, (1972), 248-259.

[6] Crowder, H. and Wolfe, P. "Linear Convergence of The conjugate gradient method", IBM Journal of Research and Development, 16, (1972), 431-433

[7] Dai, Y. H. and Yuan, Y. "A nonlinear conjugate gradient method with astrong global convergence property", SIAM J. Optim. 10 , (1999) 177-182.

[8] Fletcher, R. and Reeves, C. M. "Function minimization by conjugate gradients". Computer Journal 7 (1964) 148-154.

[9] Hager, W.W. and Zhang, H. "Algorithm 851 : CG-Descent, a conjugate gradient method with Guaranteed Descent", ACM Transactions on Mathematical Software, 32, (2006), 113-137.

[10] Hestenes, M. R. and Stiefel, E. "Methods of conjugate gradients for solving linear systems". Journal of Research of the National Bureau of Standard See., 48, (1952), 405 - 436.

[11] Polak, E. "Computational methods in optimization a unified approach". Academic Press, London, (1971).
[12] Polak, E. and Ribiere, G. "Note suria convergence de methods de directions conjuguees", RAIRO 16, (1969), 35-43.

[13] Powell, M. J. D. "Nonconvex minization calculation and the conjugate gradient method", DAMTP. 1983 INA 14, Dept. of Applied Mathematics and Theoretical Physics, University of Cambridge, England, (1983).

[14] Powell, M. J. D. "restart procedures for the conjugate gradient method, Mathematical Programming 12, (1977), 241-254.

[15] Powell, M. J. D. "Some Convergence Properties of the conjugate gradient method". Mathematical Programming 11, (1976), 42-49.

[16] Shanno, D. F. "Conjugate gradient methods with inexact searches", Mathematics of Operations Research 3, (1978), 244-256.

[17] Shanno, D.F. "Globally Convergent Conjugate Gradient Algorithms", Mathematical Programming , North-Holland 33, (1985), 61-67.

[18] Shanno, D.F. "The Convergence of a new conjugate gradient algorithm", SIAM Journal of Numerical Analysis 15, (1978), 1247 -1257.

[19] Shanno, D.F. and Phua, K. H. "Remark on algorithm 500", Transactions on Mathematical software 6, (1980), 618-622.

[20] Tassopulos, A. and Story, C. "Use of a non-quadratic model in a conjugate gradient method of optimization with inexact line search", Journal of Optimization Theory and Application, 43, (1984), 357-370.

[21] Zoutendijk, G. "Nonlinear programming, computational methods", in : J. Abadie, ed.. Integer and nonlinear programming North - Holland, Amsterdam, (1970), 37-86.
A New Restarting Criterion for FR-CG Method with…

Appendix:
All the test function used in this paper are from general literature:

1. Cubic Function (n = 2):
   \[ f = 100 \left( x_2 - x_1^3 \right)^2 + (1 - x_1)^2 \]
   \[ x_0 = (-1.2, 1)^T \]

2. Recipe Function (n = 3):
   \[ f = \left( x_1 - 5 \right)^2 + x_2^2 + x_3^2 / \left( x_2 - x_1 \right)^2 \]
   \[ x_0 = (2, 5, 1)^T \]

3. Helical Valley Function (n = 3):
   \[ f = 100 \left[ \left( x_3 - 100 \right)^2 + (4 - 1)^2 \right] + x_2^2 \]
   where
   \[ \theta = \begin{cases} 
   (2\pi)^{-1} \tan^{-1} \left( x_2 / x_1 \right) & \text{for } x_1 > 0 \\
   0.5 + (2\pi)^{-1} \tan^{-1} \left( x_2 / x_1 \right) & \text{for } x_1 < 0 
   \end{cases} \]
   \[ r = \left( x_1^2 + x_2^2 \right)^{1/2} \text{ and } x_0 = (-1, 0, 0)^T \]

4. Powell Three Variable Function (n = 3):
   \[ f = 3 - \left[ 1 / \left( 1 + (x_1 - x_2)^2 \right) \right] - \sin \left( \pi x_2 x_3 / 2 \right) - \exp \left\{- \left\{ (x_1 + x_3) / x_2 - 2 \right\}^2 \right\} \]
   \[ x_0 = (0, 1, 2)^T \]

5. Oren and Spedicato Power Function (n=10, 30, 50, 100):
   \[ f = \left( \sum_{i=1}^{n} x_i^2 \right)^2 \]
   \[ x_0 = (1; \ldots; \ldots)^T \]

6. Sum of Quadratics Function (n = 25, 70):
   \[ f = \left( \sum_{i=1}^{n} x_i - 1 \right)^T \]
   \[ x_0 = (2; \ldots; \ldots)^T \]

7. Non-Diagonal Variant of Rosenbrock Function (n = 20, 90):
   \[ f = \sum_{i=2}^{n} \left[ 100 \left( x_i - x_{i-1}^2 \right)^2 + (1 - x_i)^2 \right] ; \ n > 1 \]
   \[ x_0 = (-1; \ldots; \ldots)^T \]

8. Generalized Rosenbrock Function (n=2, 20, 60, 100):
   \[ f = \sum_{i=1}^{n/2} \left[ 100 \left( x_{2i} - x_{2i-1}^2 \right)^2 + (1 - x_{2i-1})^2 \right] \]
   \[ x_0 = (1.2, 1; \ldots; \ldots)^T \]
9. Generalized Wood Function (n=4, 20, 60, 100):

\[ f = \sum_{i=1}^{n/4} 100((x_{4i-2} - x_{4i-3})^2 + (1 - x_{4i-3})^2 + 90(x_{4i} - x_{4i-1})^2 + (1 - x_{4i-1})^2 + 10[(x_{4i-2} - 1)^2 + (x_{4i} - 1)^2] + 19.8(x_{4i-2} - 1)(x_{4i} - 1), \]

\[ x_0 = (-3, -1, -3, -1; \ldots \ldots)^T \]

10. Wolfe Function (n=80):

\[ f = \left[x_1 \left(3 - x_1/2\right) + 2x_2 - 1\right]^2 + \sum_{i=1}^{n-1} \left[x_{i-1} - x_i \left(3 - x_1/2\right) + 2x_{i+1} - 1\right]^2 + \left[x_{n-1} - x_n \left(3 - x_n/2\right) - 1\right]^2 \]

\[ x_0 = (-1; \ldots \ldots)^T \]