New Hybrid CG Algorithm Based on PR and FR Steps

Abbas Y. Al-Bayati    Khalil K. Abbo    Asma M. Abdalah
profabbasalbayati@yahoo.com
College of Computers Sciences and Mathematics
University of Mosul/Iraq

Received on: 12/11/2003    Accepted on: 09/06/2004

ABSTRACT

In this paper, a new hybrid conjugate gradient algorithm is proposed for unconstrained optimization. This algorithm combines the desirable computation aspects of Polak-Ribière steps and useful theoretical features of Fletcher-Reeves CG-steps. Computational results for this algorithm are given and compared with those of the Fletcher and Polak standard CG methods showing a considerable improvement over the latter two methods.

Keywords: unconstrained optimization, conjugate gradient algorithm, Polak-Ribière steps, Fletcher-Reeves steps.
1. Introduction

The problem of interest can be stated as that of finding a local solution \( x^* \) to the problem.

Minimize \( f(x) ; x \in \mathbb{R}^n \). \hspace{1cm} (1)

Usually \( x^* \) exists and is locally unique. There are two particular types that must be made. One is that these methods do not guarantee to find a global solution of equation (1). Another type is that the objective function \( f(x) \) must be sufficiently smooth in some cases, for more detail see (Fletcher, 1993). There are some basic theoretical results on the non-quadratic models (see Al-Bayati, 1993).

Methods for unconstrained optimization differ according to how much information the user is able to provide. The most desirable situation from the point of view of providing useful information is that the user provides subroutines from which \( f(x) \), \( g(x) \) (where \( g(x) = \nabla f(x) \)) and \( G(x) \), \( (G(x) = \nabla^2 f(x)) \) can be evaluated for any \( x \). These methods are generally iterative methods in which the user typically provides an initial estimate \( x \) of \( x^* \) and possibly some additional information. Such that at each step the search for minimum is carried out along the descent direction \( d \), i.e.

\[ d_k^T g_k < 0 \] \hspace{1cm} (2)

A sequence of iterates \( \{x_k\} \) is then generated from

\[ x_{k+1} = x_k + \lambda_k d_k \] \hspace{1cm} (3)

If the line search is exact, the step size \( \lambda_k \) is defined by

\[ \lambda_k = \arg \min_{\lambda} f(x_k + \lambda d_k) \] \hspace{1cm} (4)

In practice however an exact line search is not usually possible and any value of \( \lambda_k \) that satisfies certain standard conditions is accepted. Fletcher, (1980) suggests that \( \lambda_k \) is such that \( x_{k+1} \) satisfies the condition

\[ g_k^T d_k < - \sigma g_k^T g_k \] \hspace{1cm} (5)

together with

\[ f(x_{k+1}) \leq f(x_k) + \rho \lambda_k g_k^T d_k \] \hspace{1cm} (6)

Where \( \rho \in (0, 1/2) \), \( \sigma \in (0, 1) \) and \( \rho < \sigma \)

Conjugate gradient (CG) method is one of the few practical methods for solving large dimensionality problems because it does not require matrix storage and its iteration cost is very low. Normally the initial direction \( d_1 \) is given by

\[ d_1 = -g_1 \] \hspace{1cm} (7)

The search direction for the next iteration has the following form (see Al-Bayati and Al-Assady, 1994).

\[ d_{k+1} = -g_{k+1} + \beta_k d_k \] \hspace{1cm} (8)

Where \( \beta_k \) is a constant parameter defined either by
\[ \beta_k = \frac{\mathbf{g}_{k+1}^T \mathbf{g}_k}{\mathbf{g}_k^T \mathbf{g}_k} \ldots \ldots \] 
\[ (9a) \]

or

\[ \beta_k = \frac{\mathbf{g}_{k+1}^T (\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{g}_k^T \mathbf{g}_k} \ldots \ldots \] 
\[ (9b) \]

The definition \( \beta_k \) in (9a) is due to Fletcher and Reeves 1964 and \( \beta_k \) in (9b) is due to Polak-Ribier 1969. Extensive numerical experience has shown that the PR algorithm is more efficient than the original FR algorithm.

There is theoretical explanation which shows that PR-formula is better than FR formula. On general non-quadratic functions it can happen (see Fletcher, 1987) that the search direction \( \mathbf{d}_k \) becomes almost orthogonal to \( -\mathbf{g}_k \) and hence little progress can be made. In this event, \( x_{k+1} = x_k \) and \( g_{k+1} = g_k \) so FR method then gives

\[ \mathbf{d}_{k+1} \equiv \mathbf{d}_k \] 
\[ (10a) \]

While the PR method becomes

\[ \mathbf{d}_{k+1} \equiv -\mathbf{g}_{k+1} \] 
\[ (10b) \]

So, in this circumstance the PR algorithm tends to restart automatically to the steepest descent direction. Thus, it seems that this formula should be used when solving large problems. Many extensions and modifications have been proposed in this field (see Al-Bayati and Ahmed, 1996).

2. Theoretical results on CG methods:

Various formulas for \( \beta_k \) are given suggested in equation (8), but for purpose of this paper, attention will be focused on the (9a and 9b).

We shall assume that the level set

\[ \{x : f(x) \leq f(x_1)\} \ldots \ldots \] 
\[ (11) \]

is bounded. This assumption will ensure that \( \lambda_k \) is well defined for all \( k \). Its clear that

\[ \mathbf{d}_1^T \mathbf{g}_1 = -\mathbf{g}_1^T \mathbf{g}_1 < 0 \]

so the descent property in eq.(2) holds on the first iteration for any conjugate gradient method. Moreover, if the line search is exact, then

\[ \mathbf{g}_{k+1}^T \mathbf{d}_k = 0, \ k \geq 1 \ldots \ldots \] 
\[ (12) \]

therefore from equation (8) and (12) it follows that
This shows that a descent property holds on all iterations for any conjugate gradient method with exact line search, and in particular for both FRCG and PRCG methods.

Powell, (1983) shows that if the level set eq. (11) is bounded, if \( \lambda_k \) is defined so that eq.(12) holds for all k and if \( f(X) \) is twice continuously differentiable then FRCG method achieves the limit:

\[
\lim_{k \to \infty} \inf ||g_k|| = 0 \ldots \ldots
\]

Furthermore Al-Baali (1984) extends this result to show that even for an inexact line search satisfying (5) and (6), the descent property holds for all k and global convergence is achieved for the Fletcher-Reeves method.

Although in numerical computations (9b) is generally far more successful than formula (9a) (see Powell, 1977, 1985, for a theoretical explanation). It has not been possible to establish these global convergence results for the Polak-Ribier method unless the additional condition is imposed that the step lengths \( ||x_{k+1} - x_k|| \) tend to zero (see Powell, 1977). In fact, Powell, (1983) shows that if \( \beta_k \) is chosen to satisfy (9b) rather than (9a), then even with exact line search and exact arithmetic there exist twice continuously differentiable functions with bounded level set eq. (11) for which the gradient norms \( ||g_k||, k = 1,2, \ldots \) are bounded away from zero. This has consequently led to thoughts on how to combine the desirable computational aspects of formula (9b) and the useful theoretical features of formula (9a).

3. New hybrid CG algorithm:

In this new hybrid algorithm we assume that an inexact line search is used for non-quadratic objective function.

It can be shown that if at every iteration of the Polak-Ribier algorithm (see Story and Touti, 1990) we have

\[
g_1^T d + g_k \leq ||g_{k+1}||^2 \ldots \ldots
\]

Then the convergence proofs given by Powell (1983) and Al-Baali (1985) for Fletcher-Reeves method apply to Polak-Ribier method also-equation (15) which is an equivalent form at the equation given below is unfortunately not always satisfied

\( 0 \leq \beta_{PR} \leq \beta_{FR} \)

Consider the formula (9b)
New Hybrid CG Algorithm...

\[
\beta_{pr} = \frac{\mathbf{g}_k^T (\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{g}_k^T \mathbf{g}_k}
= \frac{\mathbf{g}_k^T \mathbf{g}_{k+1} - \mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_k^T \mathbf{g}_k}
= \frac{||\mathbf{g}_{k+1}||^2 - ||\mathbf{g}_k||^2}{||\mathbf{g}_k||^2} - \frac{||\mathbf{g}_{k+1}|| ||\mathbf{g}_k|| \cos \theta}{||\mathbf{g}_k||^2}
= \beta_{FR} - \frac{1}{\sqrt{\beta_{FR}}} \cos \theta \quad \ldots \quad (16)
\]

Where \( \theta \) is the angle between \( \mathbf{g}_{k+1} \) and \( \mathbf{g}_k \) without loss of generality suppose \( \theta \in (0, \pi/2) \)
and hence
\[
\beta_{pr} = \beta_{FR} - \frac{1}{\sqrt{\beta_{FR}}} \cos \theta \leq \beta_{FR} - \frac{1}{\sqrt{\beta_{FR}}}
\]
i.e if
\[
0 < \beta_{pr} \leq \beta_{FR} - (\beta_{FR})^{1/2} \quad \ldots \quad (17)
\]

Consequently we considered the use of hybrid conjugate gradient using formula (9b) whenever condition in eq. (17) is satisfied and formula (9a) otherwise, a descent property holds for all \( k \) and global convergence is achieved for this new hybrid algorithm when either an exact or an inexact line search is used. This algorithm was tested on several test functions and the results obtained show, in many cases, a significant improvement on the Fletcher-Reeves and Polak-Ribier methods. There were also cases, however, where the Polak algorithm performed better than this new hybrid algorithm.

4. Algorithm (New hybrid CG):

Step1: Let \( x_1 \) be an initial estimate at the minimizer \( x^* \) of \( f \)
Step2: set \( k = 1 \) and set \( d_k = -\mathbf{g}_k \)
Step3: do a line search : set \( x_{k+1} = x_k + \lambda_k d_k \)
Step4: if \( ||\mathbf{g}_{k+1}|| < \varepsilon \), where \( \varepsilon = 5 \times 10^{-5} \), take \( x^* \) as \( x_{k+1} \) and stop otherwise go to step 5
Step5: if \( k + 1 > n > 2 \) then go to step 11; otherwise, go to step 6
Step6: set the vector
\[
\tilde{d}_k = -\frac{\mathbf{g}_k^T \mathbf{d}_k}{\mathbf{g}_k^T \mathbf{g}_{k+1}} \mathbf{g}_{k+1}
\]
Step7: with
\[
\gamma = \min \left( 1, \frac{1}{\sqrt{\zeta}}, \frac{||\tilde{d}_k||}{\sqrt{||d_k||}} \right), \text{ where } \zeta
\]

31
the machine accuracy (say, $\zeta = 1 \times 10^{-16}$), we assume that
$\bar{g}_k = g(x_{k+1} - \Gamma \tilde{d}_k)$ and find

$$
\beta_{\text{FR}} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}, \quad \beta_{\text{PR}} = \frac{g_{k+1}^T (g_{k+1} - \bar{g}_k)}{g_k^T \bar{g}_k}
$$

Step 8: if $0 < \beta_{\text{PR}} \leq \beta_{\text{FR}} - \sqrt{\beta_{\text{FR}}}$ set $\beta_k = \beta_{\text{PR}}$
and go to step 9 otherwise $\beta_k = \beta_{\text{FR}}$ go to step 9
Step 9: set the search direction at the iteration $(k + 1)$
$\mathbf{d}_{k+1} = -g_{k+1} + \beta_k \bar{d}_k$
Step 10: set $k = k + 1$ and go to step 3
Step 11: set $x_{k+1} = x_1$ and go to step 2

5. Numerical Results:
All the three algorithms described in this paper, namely:
(i): The standard FRCG method
(iii): Polak-Ribier CG method
(iii): New proposed hybrid CG method
are coded in double precision Fortran 90. The numerical results are obtained
on the personal Pentium II computer. The complete set of results are given
in tables (1, 2, 3, 1a, 2a and 3a). In comparison of algorithms the number of
function evaluations (NOF) is normally assumed to be the most costly factor
in an iteration. The total number of iterations (NOI) required to achieve
convergence is also valuable in comparing similar algorithms and is also
presented here. The actual convergence criterion employed was
$g_{k+1}^T g_{k+1} < 1 \times 10^{-5}$
for all the three algorithms. Well-known test functions with different
dimensions are employed in this comparison. Tables (1a), (2a) and (3a) give
the percentage of improvements of the new proposed algorithm against FR
and PR.
Table 1: Comparison results for FR, PR and the new proposed algorithm

| Test fun. | N  | FR   |     | PR   |     | New HPF |     |
|-----------|----|------|-----|------|-----|---------|-----|
|           |    | NOI  | NOF | NOI  | NOF | NOI     | NOF |
| Cubic     | 2  | 19   | 56  | 19   | 52  | 18      | 52  |
| Rosen     | 2  | 38   | 99  | 33   | 85  | 23      | 78  |
| Powell (4)| 4  | 44   | 155 | 24   | 72  | 14      | 40  |
| Miele     | 4  | 39   | 131 | 69   | 237 | 27      | 96  |
| Dixon     | 10 | 21   | 45  | 20   | 43  | 20      | 43  |
| Cubic     | 20 | 42   | 110 | 64   | 156 | 34      | 88  |
| Wood      | 40 | 85   | 217 | 68   | 144 | 48      | 121 |
| Beal      | 40 | 42   | 87  | 10   | 26  | 30      | 68  |
| Shallow   | 60 | 47   | 97  | 53   | 107 | 26      | 57  |
| Rosen     | 60 | 122  | 271 | 92   | 231 | 63      | 145 |
| Total     |    | 499  | 1268| 442  | 1193| 303     | 988 |

Table 2: Comparison results for FR, PR and the new proposed algorithm

| Test fun. | N   | FR   |     | PR   |     | New HPF |     |
|-----------|-----|------|-----|------|-----|---------|-----|
|           |    | NOI  | NOF | NOI  | NOF | NOI     | NOF |
| Rocen     | 80  | 102  | 245 | 98   | 230 | 82      | 186 |
| Shallow   | 80  | 53   | 107 | 47   | 97  | 26      | 57  |
| Powell    | 100 | 199  | 408 | 129  | 270 | 162     | 332 |
| Wood      | 100 | 276  | 828 | 103  | 213 | 57      | 138 |
| Beal      | 120 | 47   | 95  | 12   | 28  | 30      | 59  |
| Shallow   | 120 | 49   | 101 | 53   | 109 | 29      | 63  |
| Miele     | 140 | 189  | 431 | 175  | 397 | 154     | 351 |
| Powell    | 140 | 185  | 381 | 201  | 460 | 164     | 340 |
| Rosen     | 200 | 223  | 495 | 218  | 473 | 96      | 213 |
| Beal      | 200 | 47   | 96  | 16   | 38  | 53      | 107 |
| Total     |    | 1370 | 3187| 1152 | 2315| 853     | 1846|
Table 3: Comparison results for FR, PR and the new proposed algorithm

\[ 240 \leq n \leq 600 \]

| Test fun. | N  | FR   | PR   | New HPF |
|-----------|----|------|------|---------|
|           | NOI | NOF  | NOI  | NOF     | NOI  | NOF |
| Beal      | 240 | 49   | 99   | 16      | 34   | 32   | 65   |
| Shallow   | 240 | 50   | 101  | 54      | 113  | 29   | 63   |
| Miele     | 300 | 302  | 733  | 313     | 656  | 208  | 505  |
| Rosen     | 300 | 223  | 495  | 139     | 306  | 96   | 213  |
| Shallow   | 360 | 49   | 101  | 54      | 111  | 29   | 63   |
| Wood      | 360 | 307  | 742  | 106     | 219  | 57   | 138  |
| Cubic     | 400 | 212  | 447  | 314     | 659  | 209  | 434  |
| Powell    | 400 | 346  | 708  | 413     | 877  | 405  | 825  |
| Rocen     | 600 | 332  | 720  | 282     | 591  | 96   | 213  |
| Wood      | 600 | 307  | 740  | 108     | 273  | 62   | 148  |
| Total     | 2175| 4866 | 1799 | 3839    | 1223 | 2667 |

Table (1a) Performance of the new algorithm compared with PR and FR for
\[ 2 \leq n \leq 60 \]

|          | NOI | NOF |
|----------|-----|-----|
| New      | 100%| 100%|
| FR       | 154 | 128 |
| PR       | 145 | 120 |

Table (2a) Performance of the new algorithm compared with PR and FR for
\[ 80 \leq n \leq 200 \]

|          | NOI | NOF |
|----------|-----|-----|
| New      | 100%| 100%|
| FR       | 150 | 158 |
| PR       | 135 | 125 |

Table (3a) Performance of the new algorithm compared with PR and FR for
\[ 240 \leq n \leq 600 \]

|          | NOI | NOF |
|----------|-----|-----|
| New      | 100%| 100%|
| FR       | 167 | 162 |
| PR       | 147 | 143 |
6. Conclusions:
A new proposed hybrid algorithm which combines PR and FR steps is investigated both theoretically and experimentally with obtaining a robust numerical results.

8. Appendix:

1. Generalized Sum of Quadratics Function:
\[ f(x) = \sum_{i=1}^{n} (x_i - i)^4, \quad x_0 = (2; \ldots)^T. \]

2. Generalized Osp (Oren and Spedicato Function):
\[ f(x) = \left[ \sum_{i=1}^{n} i x_i^2 \right]^2, \quad x_0 = (1; \ldots)^T. \]

3. Generalized Edger and Himmel Function:
\[ f(x) = \sum_{i=1}^{n} \left[ (x_{2i-1} - 2)^4 + (x_{2i} - 2)^2 x_{2i}^2 + (x_{2i} + 1)^2 \right], \quad x_0 = (1,0; \ldots)^T. \]

4. Generalized Canteal Function:
\[ f(x) = \sum_{i=1}^{n/2} \left[ \exp(x_{4i-3}) - x_{4i-2} \right]^4 + 100(x_{4i-2} - x_{4i-1})^6 + \right], \quad x_0 = (1,2,2,2; \ldots)^T. \]

5. Generalized Recip Function:
\[ f(x) = \sum_{i=1}^{n/3} \left[ (x_{3i-1} - 5)^2 + x_{9i-1}^2 + \frac{x_{3i}^2}{(x_{3i-1} - x_{3i-2})^2} \right], \quad x_0 = (2,5,1; \ldots)^T. \]

6. Generalized Cubic Function:
\[ f(x) = \sum_{i=1}^{n/2} \left[ 100(x_{2i} - x_{2i-1}^3)^2 + (1 - x_{2i-1})^2 \right], \quad x_0 = (-1.2,1,\ldots)^T. \]
7. Generalized Miele Function:

\[ f(x) = \sum_{i=1}^{n/4} \left[ \exp(x_{4i-3} - x_{4i-1})^2 + 100(x_{4i-2} - x_{4i-1})^6 + \right] \\
\times \left( \tan(x_{4i-1} - x_{4i}) + x_{4i}^4 + (x_{4i} - 1)^2 \right), \ x_0 = (1,2,2,2;\ldots)^T. \]

8. Generalized Dixon Function:

\[ f(x) = \sum_{i=1}^{n} \left[ (1 - x_i)^2 + (1 - x_{n})^2 + \frac{n-1}{2} \sum_{i=1}^{n} (x_i^2 - x_{i+1})^2 \right], \ x_0 = (-1;\ldots)^T. \]

9. Generalized Penalty (1) Function:

\[ f(x) = \sum_{i=1}^{n} \left[ (x_i - 1)^2 + \exp(x_i^2 - 0.25)^2 \right], \ x_0 = (1,2,\ldots,n)^T. \]

10. Generalized Penalty (2) Function:

\[ f(x) = \sum_{i=1}^{n} \left[ \exp(x_i - 1)^2 + (x_i^2 - 0.25)^2 \right], \ x_0 = (1,2,\ldots,n)^T. \]

11. Generalized Powell Function:

\[ f(x) = \sum_{i=1}^{n/3} \left[ (x_{3i-3} + 10x_{3i-2})^2 + 5(x_{3i-1} - x_{3i})^2 + \right] \\
\times \left( (x_{3i-2} - 2x_{3i-1})^4 + 10(x_{3i-3} - x_{3i})^4 \right), \ x_0 = (3,-1,0,3;\ldots)^T. \]

12. Generalized Powell 3 Function:

\[ f(x) = \sum_{i=1}^{n/3} \left[ 3 - \frac{1}{1 + (x_i - x_{i-1})^2} - \sin\left( \frac{\pi x_i x_{i-1}}{2} \right) - \exp\left[ -\left( \frac{x_i + x_{i-1}}{x_{i-1}} - 2 \right)^2 \right] \right], \ x_0 = (0,1,2;\ldots)^T. \]

13. Generalized Rosenbrock Function:

\[ f(x) = \sum_{i=1}^{n/2} \left[ 100(x_{2i} - x_{2i-1})^2 + (1 - x_{2i-1})^2 \right], \ x_0 = (-1,2,1;\ldots)^T. \]

14. Generalized Beale Function:
New Hybrid CG Algorithm…

\[ f(x) = \sum_{i=1}^{n/2} \left[ 1.5 - x_{2i-1} (1 - x_{3i}) \right]^2 + \left[ 2.25 - x_{2i-1} (1 - x_{3i}^2) \right]^2 + \left[ 2.625 - x_{2i-1} (1 - x_{3i}^3) \right]^2 \], \quad x_0 = (1, 1; \ldots)^T. \]

15. Generalized Shallow Function:

\[ f(x) = \sum_{i=1}^{n/2} x_{2i-1}^2 - x_{3i}^2 + (1 - x_{2i-1})^2, \quad x_0 = (-2, -2; \ldots)^T. \]

16. Non-Diagonal Variant of Rosenbrock Function:

\[ f(x) = \sum_{i=2}^{n} \left[ 100(x_i - x_{i-1})^2 + (1 - x_i)^2 \right], \quad n > 1, \quad x_0 = (-1; \ldots)^T. \]
REFERENCES

[1] Al-Baali, M. (1985) "Descent properly and Global Convergence of the Fletcher-Reeves Method with inexact line search", IMA Journal of Numerical Analysis, Vol. 5.

[2] Al-Bayati A. Y (1993) "A new non-quadratic model for unconstrained nonlinear optimization", J. of Mu'ta, Jordan, Vol. (8), pp. 131-155.

[3] Al-Bayati A. Y and Ahmed H. (1996) "Investigations of single update . . . . CG-methods", Qatar University Sci, Qatar, Vol. (11), pp.183-192.

[4] Al-Bayati A. Y and Al-Assady N. H (1994) "Minimization of extended quadratic function with . . . . line search", JOTA, USA, Vol. (82), pp. 139-147.

[5] Fletcher R (1980): "Practical methods at optimization", Vol. 1, Unconstrained optimization. Wiley Chichester, England 1980.

[6] Fletcher R (1987) "Practical methods at optimization", John Wiley and Sons, Chichester, New York, Brisbane, Toronto and Singapore.

[7] Fletcher R (1993): "An overview at unconstrained optimization". Numerical Analysis Report NA/149, June 1993

[8] Fletcher R and Reeves CM (1964) "Function minimization by conjugate gradient", Computer Journal 7.

[9] Polak E (1969) "Computational methods in optimization a unified approach", Academic Press, New York.

[10] Powell M. J. D (1983) "Noncenuex minimization calculations and the conjugate coradien method", Report No. DAMTP 1983/NA14, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England.

[11] Powell M. J. D (1977) "Restart procedure for the conjugate gradient method", Mathematical Programming.

[12] Powell MJD (1985) "Convergena properties of algorithms for non-linear optimization" No. DAMTP 1985/NAI. Department of Applied Mathematics and Theoretical Physics. University of Cambridge, Cambridge, England, 1985.

[13] Touati-Ahmed D and Storey (1990) "Efficient hybrid conjugate coradien techniques", Journal of Optimization Theory and Applications, Vol. 64, No. 2, 1990.