A Globally Convergence Spectral Conjugate Gradient Method for Solving Unconstrained Optimization Problems

Basim A. Hassan

basimabas39@gmail.com

College of Computer Sciences and Mathematics
University of Mosul, Mosul, Iraq

Received on: 04/10/2011 Accepted on: 14/12/2011

ABSTRACT

In this paper, a modified spectral conjugate gradient method for solving unconstrained optimization problems is studied, which has sufficient descent direction and global convergence with an inexact line searches. The Fletcher-Reeves restarting criterion was employed to the standard and new versions and gave dramatic savings in the computational time. The Numerical results show that the proposed method is effective by comparing it with the FR-method.

Keywords: Conjugate gradient method, Spectral conjugate gradient method, Numerical results.

1. Introduction

Consider the unconstrained optimization problem

\[
\min \left\{ f(x) \mid x \in \mathbb{R}^n \right\} \tag{1}
\]

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is continuously differentiable. For solution of (1), one of the algorithms in numerical performance is the Fletcher-Reeves (FR) conjugate gradient algorithm. Let \( g(x) \) denote the gradient of \( f \) at \( x \), and \( x_0 \) be an arbitrary initial approximate solution of (1). Then, in a standard FR conjugate gradient algorithm, the search direction is determined by
where

\[ \beta_k^{FR} = \frac{g_k^T g_{k+1}}{g_k^T g_k} \] (3)

Hence, a sequence of solutions will be generated by

\[ x_{k+1} = x_k + \alpha_k d_k \] (4)

where \( \alpha_k \) is the step length along \( d_{k+1} \) chosen by some kind of line search method and satisfies the strong Wolfe (SW) conditions

\[ f(x_k + \alpha k d_k) \leq f(x_k) + \delta_1 \alpha_k d_k^T g_k \] (5)

\[ |g(x_k + \alpha_k d_k)^T d_k| \leq \delta_2 d_k^T g_k \] (6)

with \( 0 < \delta_1 < \delta_2 < 1 \), where \( f_k = f(x_k) \), \( g_k = g(x_k) \), \( g_k \) is the gradient of \( f \) evaluated at the current iterate \( x_k \) \([1-4]\). In [5], Matonoha et al. (MLV) proposed another kind of conjugate gradient method, called spectral conjugate gradient method. Then, the search direction \( d_{k+1} \) in this method was defined by

\[ d_{k+1} = -\varphi_k^{MLV} g_k + \beta_k^{FR} d_k, \] (7)

where

\[ \varphi_k^{MLV} = \frac{y_k^T d_k}{g_k^T g_k} \] (8)

In this paper, we are going to develop a new conjugate gradient (CG) algorithm. The search direction generated by the method at each iteration satisfies the sufficient descent condition. We are also going to establish the global convergence of the proposed algorithm with the Wolfe-type line search.

The idea of CG methods had been studied by many researchers for example, see (Xiao et al., [6]); (Zhong et al., [7]) and (Zhang et al., [8]).

2. A New Conjugate Gradient Algorithm

If exact line search is used, the new method is identical to the MLV method. The new conjugate gradient is as follows:

\[ \varphi_k = \varphi_k^{MLV} \] as the BMLV method. Now, we present concrete algorithm as follows:

2.1. The Algorithm has the following steps:

Step 0: Given parameters \( \varepsilon = 1 * 10^{-5} \), \( \delta_1 \in (0,1) \), \( \delta_2 \in (0,1/2) \)

choose initial point \( x_0 \in R^* \).

Step 1: Computing \( g_k \); if \( \|g_k\| \leq \varepsilon \) then stop; else continue.

Step 2: Set \( d_k = -g_k \).

Step 3: Set \( \beta_k = \beta_k^{FR} \), \( \varphi_k^{BMVL} = \frac{\mu d_k^T g_{k+1} + y_k^T d_k}{g_k^T g_k} \).
Step 4: Set $x_{k+1} = x_k + \alpha_k d_k$, (Use strong Wolfe line search technique to compute the parameter $\alpha_k$)

Step 5: Compute $d_{k+1} = -g_k^{\text{BMLV}} + \beta_k d_k$.

Step 6: If $k = n$ go to step (2) with new values of $x_{k+1}$ and $g_{k+1}$. If not continue.

3. Global Convergence

In this section, we study the global convergence of Algorithm (2.1). For this, Firstly, we are going to verify that Algorithm (2.1) is well defined. For the proof of global convergence, the following assumptions 1 are needed.

**Assumption 1**

i- The level set $L = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$ is bounded.

ii- In some neighborhood $U$ and $L$, $f(x)$ is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $\mu_1 > 0$ such that

$$\|g(x_{k+1}) - g(x_k)\| \leq \mu_1 \|x_{k+1} - x_k\|, \ \forall x_{k+1}, x_k \in U.$$  

......... (10)

**Theorem (3.1)**

Suppose that $d_{k+1}$ is given by (7) and (9). Then, the following result

$${g_k^T d_{k+1}} \leq -c \|g_{k+1}\|^2 < 0$$  

......... (11)

**Proof.**

Firstly, for $k = 0$, it is easy to see that (11) is true since $d_0 = -g_0$.

Secondly, assume that

$${g_k^T d_k} \leq -c \|g_k\|^2 < 0 \text{ where } 0 < c < 1$$  

......... (12)

holds for $k$ when $k \geq 1$. Multiplying (7) by $g_{k+1}^T$, we have

$$g_{k+1}^T d_{k+1} = -g_k^{\text{BMLV}} + \beta_k g_{k+1}^T d_k$$

$$= -\mu g_k^T g_{k+1} + \nu d_k^T d_k + \beta_k g_{k+1}^T d_k$$

$$= -\mu \|g_{k+1}\|^2 \|g_k\|^2 - \|g_{k+1}\|^2 \nu d_k^T d_k + \beta_k g_{k+1}^T d_k$$

$$= -\mu d_k^T g_{k+1} \|g_{k+1}\|^2 - \|g_{k+1}\|^2 \nu d_k^T d_k + \|g_{k+1}\|^2 \beta_k g_{k+1}^T d_k$$

$$= -\mu d_k^T g_{k+1} \|g_{k+1}\|^2 - c \|g_{k+1}\|^2 + \mu \|d_k g_{k+1}\|^2 \beta_k$$

$$= -\|g_{k+1}\|^2 \left( c + \frac{\mu d_k^T g_{k+1}}{\|g_k\|^2} \right)$$

......... (13)
from (6) and (12), we get
\[
g^T_{k+1}d_{k+1} \leq -\|g_{k+1}\|^2 \left[ c + \frac{\mu_\delta_c \|g_k\|^2}{\|g_k\|^2} \right] \\
\leq -\|g_{k+1}\|^2 \left[ c + \mu_\delta_c \right] \\
\leq -c_1\|g_{k+1}\|^2
\]

where \( c_1 = c + \mu_\delta_c \) is positive constant.

**Theorem (3.2)**

Consider the conjugate gradient algorithm 2.1 where \( \varphi_k \) and \( \beta_k \) are given by (8) and (3) respectively and \( \alpha_k \) is obtained by the strong Wolfe line search (5) and (6). Suppose that the assumptions (i) and (ii) hold. Then either

\[
\lim \inf_{k \to \infty} \|g_k\| = 0 \text{ or } \sum_{k=1}^\infty (g^T_k d_k)^2 < +\infty.
\]

**Proof:**

If our conclusion does not hold, then there exists a real number \( \varepsilon_i > 0 \) such that \( \|g_{k+1}\| > \varepsilon_i \) for all \( k = 1, 2, 3, \ldots \). Squaring the both terms of \( d_{k+1} + \varphi_k BMLV g_{k+1} = \beta_k d_k \), we get

\[
\|d_{k+1}\|^2 + (\varphi_k BMLV \|g_{k+1}\|^2 + 2\varphi_k BMLV d^T_{k+1}g_{k+1} = \beta_k^2 \|d_k\|^2
\]

from (16), we get

\[
\|d_{k+1}\|^2 = \beta_k^2 \|d_k\|^2 - 2\varphi_k BMLV d^T_{k+1}g_{k+1} - (\varphi_k BMLV)^2 \|g_{k+1}\|^2
\]

Dividing both sides of (17) by \( (g^T_k d_k)^2 \), by (3), (11) and \( \|g_{k+1}\| > \varepsilon_i \), we have

\[
\frac{\|d_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} = \frac{\|g_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} - (\varphi_k BMLV)^2 \frac{\|g_{k+1}\|^2}{(d^T_{k+1}g_{k+1})^2} - 2\varphi_k BMLV \frac{1}{d^T_{k+1}g_{k+1}}
\]

\[
\leq \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2} - (\varphi_k BMLV)^2 \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2} - 2\varphi_k BMLV \frac{1}{\|g_{k+1}\|^2}
\]

\[
\leq \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2} - (\varphi_k BMLV)^2 \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2} - 2\varphi_k BMLV \frac{1}{\|g_{k+1}\|^2} + 1
\]

\[
\|d_{k+1}\|^2 \leq d^T_{k+1}g_{k+1} + 1 \leq \|d_k\|^2 + \frac{1}{\|g_{k+1}\|^2}
\]

Since \( d_k = -g_k \), so that
\[
\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_k + e_i^2)} < \frac{\|d_k\|^2}{(d_k^T g_k + e_i^2)} + \frac{k - 1}{e_i^2} \frac{1}{g_k} \frac{k - 1}{e_i^2} + \frac{k - 1}{e_i^2} = \frac{k}{e_i^2} \quad \text{......... (20)}
\]

Thus,
\[
\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} > \sum_{k=1}^{\infty} \frac{e_i^2}{k} = \infty \quad \text{......... (21)}
\]

Which is contrary to Theorem (3.2). The proof is complete.

4. Numerical Results

In this section, we reported some numerical results obtained with the implementation of the new algorithm on a set of unconstrained optimization test problems. We have selected (9) large scale unconstrained optimization problems in extended or generalized form, for each test function, we have considered numerical experiment with the number of variable n=100-1000. Using the standard Wolfe line search conditions (4) and (5) with \(\delta_1 = 0.0001\) and \(\delta_2 = 0.9\) In all of these cases, the stopping criteria is the \(\|g_k\| \leq 10^{-5}\). The programs were written in Fortran 90. The test functions were commonly used for unconstrained test problems with standard starting points and a summary of the results of these test functions was given in Table (3.1). We tabulate for comparison of these algorithms, the number of function evaluations (NOF) and the number of iterations (NOI).

| No. | n     | FR-algorithm | New-algorithm | MLV-algorithm |
|-----|-------|--------------|---------------|---------------|
|     |       | NOF (NOI)    | NOF (NOI)     | NOF (NOI)     |
| 1   | 100   | 872 (323)    | 228 (112)     | 774 (304)     |
|     | 1000  | 7741 (2005)  | 706 (351)     | Failed        |
|     |       |              |               |               |
| 2   | 100   | 242 (119)    | 222 (108)     | 211 (104)     |
|     | 1000  | 1272 (634)   | Failed        | 369 (183)     |
|     |       |              |               |               |
| 3   | 100   | 209 (102)    | 109 (53)      | 209 (102)     |
|     | 1000  | 563 (279)    | 45 (21)       | 561 (278)     |
|     |       |              |               |               |
| 4   | 100   | 204 (31)     | 130 (34)      | 230 (34)      |
|     | 1000  | 264 (35)     | 105 (45)      | 283 (38)      |
|     |       |              |               |               |
| 5   | 100   | 297 (103)    | 152 (72)      | 280 (107)     |
|     | 1000  | 408 (159)    | 366 (176)     | 598 (267)     |
|     |       |              |               |               |
| 6   | 100   | 271 (121)    | 266 (123)     | 303 (119)     |
|     | 1000  | 2253 (1001)  | 942 (1886)    | 845 (2128)    |
|     |       |              |               |               |
| 7   | 100   | 209          | 412           | 218           |
5. Conclusions and Discussions.

In this paper, we have proposed modified spectral CG method for solving unconstrained minimization problems. The computational experiments show that the new approaches given in this paper are successful.

Table (4.1) gives a comparison between the new-algorithm and the Fletcher-Reeves (FR)-algorithm for convex optimization; this table indicates that the new algorithm and MLV-algorithm save $(76.50–96.14\%)$ NOI and $(65.61–98.94\%)$ NOF, overall against the standard Fletcher-Reeves (FR)-algorithm, especially for our selected test problems.

Relative Efficiency of the Different Methods Discussed in the Paper.

| Tools           | NOI   | NOF   |
|-----------------|-------|-------|
| FR -algorithm   | 100 % | 100 % |
| MLV-algorithm   | 3.85 %| 1.05 %|
| New-algorithm   | 23.49 %| 34.38 %|

APPENDIX

1. Generalized wood function:

\[
f(x) = \sum_{i=1}^{\infty} 4(x_{4i-2} - x_{4i-3}^2) + (1 - x_{4i-3})^2 + 90(x_{4i} - x_{4i-1}^2)^2 + (1 - x_{4i-1})^2 + 10.1((x_{4i-2} - 1)^2 + (x_{4i} - 1)^2) + 19.8((x_{4i-1} - 2) + (x_{4i} - 1))
\]

Starting point: $(-3,-1,-3,-1,\ldots)$

2. Helical valley function:

\[f(x) = 100((x_1 - 10\theta)^2 + (r - 1)^2) + x_3^2 \quad \text{where} \quad \theta = \begin{cases} (2\Pi)^{-1} \tan(x_1 / x_3) & \text{for} \; x_1 > 0 \\ 0.5 + (2\Pi)^{-1} \tan(x_2 / x_3) & \text{for} \; x_1 < 0 \end{cases}
\]

\[r = (x_1^2 + x_3^2)^{1/2}
\]

Starting point: $(-1,0,0,\ldots)$


3. Penalty function:
\[ f(x) = \sum_{i=1}^{n} e^{(x(i)-1)^2} + (x(i)^2 - 0.25)^2 \]

**Starting point**: (1,2, ................. ...)^T

4. Cantrell function:
\[ f(x) = \sum_{i=1}^{n/4} [\exp(x_{4i-3}) - x_{4i-3}]^4 + 100(x_{4i-2} - x_{4i-1})^6 + [\tan^{-1}(x_{4i-1} - x_{4i})]^4 + x_{4i-3}^8 \]

**Starting point**: (1, 2, 2, 2, ................. ...)^T

5. Rosenbrock function:
\[ f(x) = \sum_{i=1}^{n/2} (100(x_{2i} - x_{2i-1})^2 + (1 - x_{2i-1})^2) \]

**Starting point**: (−1,2,1,−1,2,1,......)^T

6. Miele function:
\[ f(x) = \sum_{i=1}^{n/4} [\exp(x_{4i-3}) - x_{4i-3}]^3 + 100(x_{4i-2} - x_{4i-1})^6 + [\tan(x_{4i-1} - x_{4i})]^4 + x_{4i-3}^8 + (x_i - 1)^3 \]

**Starting point**: (1, 2, 2, 2, ................. ...)^T

7. Non-diagonal function:
\[ f(x) = \sum_{i=1}^{n/2} (100(x_i - x_i^3)^2 + (1 - x_i)^2) \]

**Starting point**: (−1, ................. ........... ...)^T

8. Welfe function:
\[ f(x) = (-x_i(3-x_i/2)+2x_i-1)^2 + \sum_{i=1}^{n/4} (x_{4i-1} - x_i(3-x_i(3-x_i/2)+2x_i-1)^2 + (x_{4i-1} - x_i(3x_i/2-1))^3 \]

**Starting point**: (−1, ................. ........... ...)^T

9. Sum of Quartics function:
\[ f(x) = \sum_{i=1}^{n} (x_i - 1)^4 \]

**Starting point**: (2, ................. ........... ...)^T
REFERENCES

[1] Andrie, N. (2008) 'An Unconstrained Optimization Test functions collection' Advanced Modeling and optimization. 10, pp.147-161.

[2] Dai, Y. and Yuan Y. (1999) 'A nonlinear conjugate gradients method with a strong global convergence property' SIAM J. Optimization 10, pp. 177-182.

[3] Fletcher R. (1987) 'Practical Methods of Optimization' (second edition). John Wiley and Sons, New York.

[4] Fletcher, R. and Reeves C. (1964) 'Function minimization by conjugate gradients' Computer J. 7, pp.149-154.

[5] Matonoha C., Luksan L. and Vlcek J. (2008) 'Computational experience with conjugate gradient methods for unconstrained optimization' Technical report no.1038, pp.1-17.

[6] Xiao Y., Wang Q. and Wang D. (2010) 'Note on the Dai-Yuan-Yuan modified spectral gradient method' J. of Computational and Applied Math., 234, pp.2986-2992.

[7] Zhong W., Zan L.Y. and YaLin W. (2011) 'New spectral PRP conjugate gradient method for unconstrained optimization' Applied Math. Letters, 244, pp.16-22.

[8] Zhang L., Zhou W. Li D. (2006) 'Global convergence of a modified Fletcher-Reeves conjugate gradient method New spectral PRP conjugate gradient method With Armijo-type line search' Number Math., 104, pp.561-572.