Available online at http://scik.org

J. Math. Comput. Sci. 2 (2012), No. 5, 1363-1376

ISSN: 1927-5307

# SCALED SEARCH DIRECTIONS FOR NON-LINEAR CONJUGATE GRADIENT METHODS

ABBAS Y. Al-BAYATI[1,*], OMAR M. Al-SHAHER[2]

[1]College of Basic Education Telafer, Mosul University, Mosul, Iraq

[2]College of Computers Sciences and Mathematics, Mosul University, Mosul, Iraq

**Abstract:** In this paper, we have suggested an acceleration step-length for the search directions of a new proposed minimization algorithm, which is based on function and gradient values, we have dealt with the global convergent property for the new suggested algorithm and we have obtained sufficiently good numerical results by implementing some newly written Fortran programs in this field.

**Keywords:** Unconstrained Optimization, Conjugate Gradient, Sufficient Descent, Acceleration Step-Length, Global Convergence.

**2000 AMS Subject Classification:** 47H17; 47H05; 47H09

## 1. Introduction

The Conjugate Gradient (CG) methods are useful in finding the minimum value of a function with large number of variables for unconstrained minimization problems. Generally, for (n) number of variables the method has the following from:

$$\min f(x);\ x \in R^n \qquad (1)$$

[*]Corresponding author

Received April 21, 2012

where $f:R^n \to R$ is continuously differentiable. The CG-method is an iterative method of the form:

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

where $x_k$ is the current iterate point, $\alpha_k > 0$ is a step-length and $d_k$ is the search direction. The search direction $d_k$ is defined by:

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

In (3) $\beta_k$ is known as the CG parameter, the line search in the CG-method often is based on the Wolfe conditions:

$$\left.\begin{aligned} &f(x_{k-1} + \alpha_k d_k) - f(x_{k-1}) \leq \delta \alpha_k g_{k-1}^T d_{k-1}, \\ &g_k^T d_{k-1} \geq \sigma g_{k-1}^T d_{k-1}, \end{aligned}\right\} \tag{4}$$

$$\left.\begin{aligned} &f(x_{k-1} + \alpha_k d_k) - f(x_{k-1}) \leq \delta \alpha_k g_{k-1}^T d_{k-1}, \\ &\left|g_k^T d_{k-1}\right| < -\sigma g_{k-1}^T d_{k-1} \end{aligned}\right\} \tag{5}$$

where $d_k$ is a descent direction and $0 < \delta \leq \sigma < 1$. Equations (4) and (5) are called the standard and strong Wolfe conditions respectively. Different CG-methods correspond to different choices for the parameter $\beta_k$. The parameter $\beta_k$ in (3) is selected so that when applied to minimize a strongly quadratic convex function, the direction $d_k$ is conjugate subject to the Hessian of the quadratic function. Some well-known formulas are given as follows:

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \tag{6}$$

$$\beta_k^{BA} = \frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}} \tag{7}$$

$$\beta_k^{PR} = \frac{g_k^T y_{k-1}}{\|g_{k=1}\|^2} \tag{8}$$

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} \tag{9}$$

$$\beta_k^{LS} = \frac{g_k^T (g_k - g_{k-1})}{-d_{k-1} g_{k-1}} \tag{10}$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{y_{k-1}^T d_{k-1}} \tag{11}$$

$$\beta_k^{CD} = -\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} \tag{12}$$

$$\beta_k^{DL} = \frac{g_k^T (y_{k-1} - t s_{k-1})}{d_{k-1}^T y_{k-1}} \tag{13}$$

where $y_{k-1} = g_k - g_{k-1}$ and $s_{k-1} = x_k - x_{k-1} = \alpha_{k-1} d_{k-1}$, with a parameter $t \in [0, \infty)$. Here, $g_k$ and $g_{k-1}$ are the gradients of $f(x)$ at the point $x_k$ and $x_{k-1}$ respectively. For the above corresponding methods, FR is known as Fletcher and Reeves [7], BA is known as Al-Bayati and Al-Assady [1], PR is known as Polak and Ribiere [14], HS is known as Hestenes and Stiefel [11], LS is known as Liu and Storey [12], DY is known as Dai and Yuan [5], CD is known as Conjugate Descent by Fletcher [8] and lastly DL is known as Dai and Liao [4]. Dai and Yuan [6] and Yuan and Sun [19] have shown that, all these methods are equivalent for $f(x)$ that is strictly convex quadratic function, but behaves differently for general non quadratic functions. An important feature of the HS method is that it satisfies conjugacy condition:

$$d_k^T y_{k-1} = 0, \tag{14}$$

which is independent of the objective function and line search. However, Dai and Liao [4] pointed out that in the case $g_{k+1}^T d_k \neq 0$, the conjugacy condition (14) may have some disadvantages. In order to construct a better formula for $\beta_k$, Dai and Liao proposed a CG-method with a new conjugacy condition called DL-method, given by :

$$\beta_k^{DL} = \frac{g_k^T (y_{k-1} - t s_{k-1})}{d_{k-1}^T y_{k-1}} \tag{15}$$

Based on the idea of the DL-method, Hager and Zhang (HZ) [10] proposed a descent CG-method. Besides CG-method, the following gradient type methods :

$$d_k = \begin{cases} -g_k & if \;\; k = 0, \\ -\theta_k g_k + \beta_k d_{k-1}, & if \;\; k > 0, \end{cases} \tag{16}$$

Here $\theta_k$ and $\beta_k$ are two parameters. Clearly, if $\theta_k = 1$, the methods (13) become standard CG-methods (3). Also, when:

$$\beta_k = \frac{(\theta_k y_{k-1} - s_{k-1})^T g_k}{d_{k-1}^T y_{k-1}}, \tag{17}$$

Zhang, Zhou and Li [18] proposed a modified FR method where the parameters in (15) are given by:

$$\theta_k = \frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2}, \ \beta_k = \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \ . \tag{18}$$

This method satisfies $g_{k-1}^T d_{k-1} = -\|g_{k-1}\|^2$ and moreover, this method converges globally for non-convex functions with Armijo or Wolfe line search.

In this paper, we are concerned with the CG-method defined as in (16). However, in the next section, we present an acceleration step-length for scaling the search directions**.** In section 3, we have analyzed the global convergence properties of the new proposed algorithm. In section 4, we have reported some numerical comparisons against Zhang's CG-method defined in [19] using (35) different test problems from the CUTE [3].

## 2. Materials and Methods.

In this section we have described the following two-terms PRCG type method for any spectral parameter $\theta_k$ defined by:

$$d_k = \begin{cases} -g_k & if\, k = 0 \\ -\theta g_k + \beta_k^{PR} d_{k-1}, & if\, k > 0 \end{cases} \tag{19}$$

### 2.1. Zhang's PRCG-Method [19].

We have used the following search direction to calculate Zhang's CG-method by:

$$d_k = \begin{cases} -g_k & if\, k = 0 \\ -\theta_k^{PRP} g_k + \beta_k^{PRP} d_{k-1}, & if\, k > 0 \end{cases} \tag{20}$$

$$\left.\theta_k^{PRP} = 1 + \beta_k^{PRP} \frac{g_k^T d_{k-1}}{\|g_k\|^2} - \psi \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} \;\; ; \;\; \beta_k^{PRP} = \frac{g_k^T z_{k-1}}{\|g_{k-1}\|^2} \right\} \tag{21}$$

**2.2. The Acceleration Step-length.**

In CG-method (see Nocedal [13]) the search directions tend to poorly scaled and as a consequence the line search must perform more function evaluation in order to obtain a suitable step-length $\alpha_k$ .

As in Andrei's [2] an accelerated scheme for the standard Wolfe line search procedure was used by proceeding as follows :

$$x_{k+1} = x_k + \xi_k \alpha_k d_k \,, \tag{22}$$

$$\xi_k = -\frac{a_k}{b_k} \,,\; a_k = \alpha_k g_k^T d_k \;\;,\; b_k = -\alpha_k (g_k - g_z)^T d_k \;,\; g_z = \nabla f(z) \tag{23}$$

$$z = x_k + \alpha_k d_k \tag{24}$$

If $b_k \neq 0$, compute $x_{k+1} = x_k + \xi_k \alpha_k d_k$ ; else, compute

$$x_{k+1} = x_k + \alpha_k d_k \,, \tag{25}$$

**2.3. New Acceleration Search Directions for the New Proposed Algorithm.**

We design our new search directions in our new proposed algorithm as follows:

$$d_k^{scaled} = -\psi_k \frac{\|g_k\|}{\|d_k\|} sign(d_k^T g_k) d_k \,, \tag{26}$$

where $d_k$ follows (3) and $\beta_k = \beta_k^{PR}$ in (3), $\psi_k > 0$ is a bounded scalar, and $sign(g_k^T d_k)$ is the sign function defined by

$$sign(t) = \begin{cases} 1, & if\, t > 0 \\ -1 & if\, t < 0. \end{cases} \tag{27}$$

If $d_k^T g_k = 0$, we design $d_k^{scaled} = -g_k$ as the restart condition.

**2.4. Outline of the New Scaled CG-Algorithm.**

**Step(0):** Given $\psi \in [0,1]$; $k=0$, choose an initial point $x_0 \in R^n, \varepsilon > 0$.

**Step(1):** Set $d_k = -g_k$.

**Step(2):** Compute $\alpha_k$ by using (4)-(5)-(22)-(23)-(24) and (25).

**Step(3):** Compute $f_k$ and $g_k$ ; $s_{k-1} = x_k - x_{k-1}$.

**Step(4):** If $\|g_k\| \le \varepsilon$ stop .

**Step(5):** If $\left|g_k^T g_{k-1}\right| \ge 0.2\|g_k\|^2$, go to **Step(1)** else continue.

**Step(6):** Compute Zhang's search direction:

$$\left.\begin{aligned} d_k &= -\theta_k^{PRP} g_k + \beta_k^{PRP} d_{k-1} \\ \theta_k^{PRP} &= 1 + \beta_k^{PRP} \frac{g_k^T d_{k-1}}{\|g_k\|^2} - \psi \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} \\ \beta_k^{PRP} &= \frac{g_k^T z_{k-1}}{\|g_{k-1}\|^2} \end{aligned}\right\} \tag{28}$$

**Step(7)**: Compute the new scaled search directions by setting [16]:

(i) Set p=1 and $d_k^{scaled} = -\psi_k \dfrac{\|g_k\|}{\|d_k\|} sign(d_k^T g_k) d_k$ (29a)

(ii) If $g_k^T d_k^{scaled} = 0$ $\quad then\, put \quad d_k^{scaled} = -g_k$

(iii) If $g_k^T d_k^{scaled} > -c^p \|g_k\|^2$; set $p = p+1$ and repeat. (29b)

(iv) Let $d_k = d_k^{scaled}$, put $k = k+1$ and go to **Step(2)**

## 3. Convergence properties.

In this section, we only analyze the convergence properties of the new proposed algorithm. In the global convergence analysis of many iterative methods, the following assumption is often needed. The following theorem is often used to prove global convergence of

CG-methods. It was originally given by Zoutendijk [20] and Wolfe [15].

**Assumption 3.1.**

(i) The level set $\Omega = \{x \in R^n : f(x) \le f(x_0)\}$ is bounded .

(ii) In some neighborhood $N$ of $\Omega$ , $f$ is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that :

$$\|g(x) - g(y)\| \le L\|x - y\|, \qquad \forall x, y \in N \tag{30}$$

Clearly, Assumption (i) implies that there exists a constant $\gamma$ such that:

$$\|g(x)\| \le \gamma \text{ for all } x \in N \tag{31}$$

**Theorem 3.2.** Let $\{x_k\}$ and $\{d_k\}$ be generated by the new proposed algorithm, and let $\alpha_k$ be obtained by the modified Wolfe line search procedure defined by (4)-(5)-(22)-(23)-(24) and (25). If $\psi \in [0,1]$, then we have:

$$\frac{g_k^T d_k}{\|g_k\|^2} \le -(1-\psi) \tag{32}$$

**Proof.**

$$d_k = -\theta_k^{PRP} g_k + \beta_k^{PRP} d_{k-1}$$

$$\theta_k^{PRP} = 1 + \beta_k^{PRP} \frac{g_k^T d_{k-1}}{\|g_k\|^2} - \psi \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}}$$

$$d_k = -g_k - \beta_k^{PRP} \frac{\|g_k\|^2 d_{k-1}}{\|g_k\|^2} + \psi \frac{\|g_k\|^2 d_{k-1}}{d_{k-1}^T y_{k-1}} + \beta_k^{PRP} d_{k-1}$$

$$g_k^T d_k = -\|g_k\|^2 - \beta_k^{PRP} g_k^T d_{k-1} + \psi \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} + \beta_k^{PRP} g_k^T d_{k-1}$$

$$g_k^T d_k = -\|g_k\|^2 + \psi \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} \tag{33}$$

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

$$\frac{g_k^T d_k}{\|g_k\|^2} = -1 + \psi \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} \tag{34}$$

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

If use ELS then we get: $\frac{g_k^T d_k}{\|g_k\|^2} = -(1-\psi)$

We proceed by induction**,** initially, for k=0, we have: $g_0^T d_0 = -\|g_0\|^2 < 0.$

Assume for all (k-1) we have: $g_{k-1}^T d_{k-1} = -\|g_{k-1}\|^2 < 0$

Now for general k, we have to show that: $\frac{g_k^T d_k}{\|g_k\|^2} \le -c\,;\; c > 0$

Now, we have: $\frac{g_k^T d_k}{\|g_k\|^2} \le -(1-\psi)$ The proof is then finished.

**Theorem 3.3.**

Let Assumption-3.1 hold. Consider any iteration method of the form (2)-(3), where $d_k$ satisfy $g_k^T d_k < 0$, If $\alpha_k$ satisfies the Wolfe condition (4) or the Strong Wolfe condition (5) , then we have:

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)}{\|d_k\|^2} < +\infty. \tag{35}$$

Since the new algorithm has the sufficient descent property and since it uses the strong Wolfe condition (5) then the new proposed algorithm satisfies the global convergence property, we can back to a basic theorem proved in the [9].

## 4. Numerical Results.

The main work of this section is to report the performance of the new proposed algorithm on a set of test problems. The codes are newly written in Fortran and in double precision arithmetic. All the tests are performed on a PC. Our experiments are performed on a set of (35) nonlinear unconstrained problems that have second derivatives available. These test problems are contributed in CUTE [3] and their details are given in the Appendix. For each test function we have considered 10 numerical experiments with number of variables n = 100, 200, . . . ,1000. In order to assess the reliability of our new proposed method, we have tested them against Zhang's [19] method using the same test problems. All these methods terminate when the following stopping criterion is met:

$$\|g_k\| \leq 10^{-6}. \quad (36)$$

We also force these routines stopped if the iterations exceed 1000 or the number of function evaluation reach 2000 without achieving the minimum. We use $\delta = 10^{-4}$, $\sigma = 0.1$ , in some part of the modified line search routine used in this new method.

**Table (4.1)** compares some numerical result for new method respectively against Zhang's method; these table use $\varepsilon_1 = 0.1$ and $\psi = 0.5$ and they indicate for **(n)** as a dimension of the problem; **(NOI)**, number of iterations; **(NOFG)**, number of function and gradient evaluations; **(TIME)**, the total time required to complete the evaluation process for each test problem.

In **Table (4.2)** we have compared the percentage performance of the new and Zhang's methods taking over all the tools as 100% . In order to summarize our numerical results, we have concerned only on the total of different dimensions n=100, 200. . . 1000 for all Tools used in these comparisons.

**Table (4.1)**

**Comparisons between the New and Zhang's (2007) CG-method**

**for n=100, 200. . . 1000**

| **Prob.** | **Zhang (2007).** NOI /NOFG/TIME | **New Method.** NOI/NOFG/TIME |
|---|---|---|
| **1** | 2366/2744/1.63 | 825/1120/0.47 |
| **2** | 301/739/0.09 | 215/423/0.02 |
| **3** | 126/385/0.12 | 75/96/0.03 |

| | | |
|---|---|---|
| **4** | 909/1676/0.92 | 696/844/0.28 |
| **5** | 376/721/0.12 | 506/595/0.07 |
| **6** | 316/588/0.41 | 319/346/0.20 |
| **7** | 582/1193/0.12 | 1279/1307/0.11 |
| **8** | 92/343/0.28 | 161/196/0.14 |
| **9** | 246/593/0.07 | 234/365/0.01 |
| **10** | 211/508/0.23 | 197/241/0.08 |
| **11** | 442/872/0.22 | 603/724/0.12 |
| **12** | 207/531/0.06 | 195/313/0.04 |
| **13** | 70/344/0.06 | 32/64/0.00 |
| **14** | 797/1487/0.23 | 645/728/0.06 |
| **15** | 1176/2117/0.38 | 1024/1095/0.13 |
| **16** | 156/422/0.12 | 93/116/0.01 |
| **17** | 160/408/0.14 | 90/118/0.02 |
| **18** | 153/413/0.14 | 109/133/0.00 |
| **19** | 393/808/0.16 | 614/703/0.08 |
| **20** | 125/383/0.11 | 75/96/0.01 |
| **21** | 416/873/0.13 | 608/701/0.06 |
| **22** | 390/693/0.26 | 504/579/0.14 |
| **23** | 830/1593/0.66 | 739/806/0.23 |
| **24** | 166/452/0.05 | 124/207/0.02 |
| **25** | 253/608/0.09 | 236/267/0.03 |
| **26** | 2730/3403/2.13 | 955/1039/0.25 |
| **27** | 148/408/0.05 | 137/189/0.01 |
| **28** | 131/403/0.15 | 80/160/0.03 |
| **29** | 122/376/0.22 | 85/105/0.05 |
| **30** | 128/471/0.19 | 44/76/0.01 |
| **31** | 211/508/0.24 | 198/241/0.06 |
| **32** | 730/1393/0.18 | 554/677/0.07 |

| | | |
|---|---|---|
| **33** | 10/30/0.01 | 29/84/0.00 |
| **34** | 90/110/0.05 | 80/110/0.02 |
| **35** | 233/536/0.05 | 90/310/0.01 |
| **Total** | **15792/29132/10.07** | **12450/15174/2.87** |

**TABLE (4.2)**

**Percentage performance of the New Method against Zhang's (23007) CG-method**

| TOOLS | Zhang (2007) | New Method |
|---|---|---|
| **NOI** | 100% | 78.9% |
| **NOFG** | 100% | 52.1% |
| **TIME** | 100% | 28.6% |

## 5. Discussion.

It is clear from **Table (4.2)** that taking, over all, the Tools as a 100% for the Zhang's method, the New method has an improvement, in about (21.1)% NOI; (47.9)% NOFG and (71.4)% TIME, these results indicate that New method is in general is the best .

**APPENDIX.**

The details of the test functions, used in this paper, can be found in CUTE [3]. The numbers (1-35) in our tables indicate to:

1- Extended Trigonometric Function.
2- Extended Penalty Function.
3- Raydan 2 Function.
4- Extended Hager Function.
5- Generalized Tridiagonal-1 Function.
6- Extended 3-Exponential Terms Function.
7- Diagonal4 Function.
8- Diagonal5 Function.

9- Extended Himmelblau Function.

10- Extended PSC1 function.

11- Extended Block Diagonal BD1 function.

12- Extended Quadratic Penalty QP1.

13- Extended EP1 Function.

14- Extended Tri-diagonal 2 Function.

15- ARWHEAD (CUTE)-Function.

16- DIXMAANA (CUTE)-Function.

17- DIXMAANB (CUTE)-Function.

18- DIXMAANC (CUTE)-Function.

19- EDENSCH (CUTE)-Function.

20- DIAGONAL 6 Function.

21- ENGVALI CUTE-Function.

22- DENSCHNA CUTE-Function.

23- DENSCHNC (CUTE)-Function.

24- DENSCHNB (CUTE)-Function.

25- DENSCHNF (CUTE)-Function.

26- Extended Block-Diagonal BD2 Function.

27- Generalized quadratic GQ1 Function.

28- DIAGONAL 7 Function.

29- DIAGONAL 8 Function.

30- Full Hessian Function.

31- SINCOS Function.

32- Generalized quadratic GQ2 Function.

33- ARGLINB (CUTE)-Function.

34- HIMMELBG CUTE-Function.

35- HIMMELBH CUTE-Function.

REFERENCES

[1] A. Y. AL-Bayati and N. H. AL-Assady, Conjugate gradient methods, Technical Research Report, No (1/86), School of Computer Studies, Leeds University, U.K., (1986).

[2] N. Andrei, Acceleration of conjugate gradient algorithms for unconstrained optimization, Applied Mathematics and Computation., 213, (2009), 361-369.

[3] K. E. Bongartz; A. R. Conn, N.I.M. Gould and P.L. Toint, CUTE: constrained and unconstrained testing environments. ACM Trans. Math. Softw., 21, (1995), 123-160.

[4] Y. H. Dai and L. Z. Liao, New conjugate conditions and related nonlinear conjugate gradient methods, Appl. Math. Optim., 43, (2001), 87-101.

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

[6] Y. H. Dai and Y. Yuan, Nonlinear conjugate gradient method, Beijing: Shanghai Scientific and Technical Publishers, (1998).

[7] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients, Comp. J., 7, (1964), 149-154.

[8] R. Fletcher, "Practical Methods of Optimization", Part l: unconstrained Optimization, John Wiley & Sons, New York (1987).

[9] J. C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optim.,2(1), (1992), 2-42.

[10] W. W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM J. Optim., 16, (2005), 170-192.

[11] M. R. Hestenes and E. Stiefel, Method of conjugate gradient for solving liner equations, J. Res. Nat. Bur. Stand., 49, (1952), 409-436.

[12] Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms part 1: theory, J. Comput. App1. Math., (1992), 129-137.

[13] J. Nocedal, Conjugate gradient methods and nonlinear optimization. In Linear and nonlinear Conjugate Gradient related methods, Adams L. and Nazareth J.L. (eds.), SIAM., 9-23 (1996).

[14] E. Polak and G. Ribiere, Note sur la convergence de directions conjugues. Rev. Francaise Inform. Recherche Opérationnelle, 16, (1969), 35-43.

[15] P. Wolfe, Convergence conditions for ascent methods, SIAM Rev., 11, (1969), 226-235.

[16] G. Yuan and X. Lu, A modified PRP conjugate gradient method. Ann. Oper. Res., 166, (2009), 73-90.

[17] Y. Yuan and W. Sun, Theory and methods of optimization, Beijing, Science Press of China, (1999).

[18] L. Zhang; W. Zhou and D. Li, Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search. Number Math., 104, (2006), 561-572.

[19] L. Zhang and D. Li, New versions of the Hestenes-Stiefel nonlinear conjugate gradient method based on the secant condition for optimization, Computational and Applied Mathematics., 28, (2009), 111-133.

[20] G. Zoutendijk, "Nonlinear Programming, Computational methods, Integer and Nonlinear Programming". J. Abadie (ed.), North-Holland, Amsterdam, (1970), 37-86.