Communications in Applied Analysis, **21**, No. 1 (2017), 15-28 ISSN: 1083-2564

GLOBAL CONVERGENCE OF THE TMR METHOD FOR UNCONSTRAINED OPTIMIZATION PROBLEMS

T. BOUALI[1], M. BELLOUFI[2], AND R. GUEFAIFIA[3]

[1,3]Departement of Mathematics and Informatics
University Tebessa
ALGERIA

[2]Laboratory Informatics and Mathematics (LiM)
Faculty of Science and Technology
Department of Mathematics and Informatics
Mohamed Cherif Messaadia University
Souk Ahras, ALGERIA

ABSTRACT: Conjugate gradient methods are probably the most famous iterative methods for solving large scale optimization problems in scientific and engineering computation, characterized by the simplicity of their iteration and their low memory requirements. It is well known that the search direction plays a main role in the line search method. In this paper, we propose a new search direction with the Wolfe line search technique for solving unconstrained optimization problems. Under the above line searches and some assumptions, the global convergence properties of the given methods are discussed. Numerical result shows that the proposed formula is superior and more efficient when compared to other CG coefficients.

AMS Subject Classification: 90C26, 65H10

Key Words: unconstrained optimization, conjugate gradient method, strong wolfe line search, global convergence

Received: October 12, 2016 ; **Accepted:** November 30, 2016 ;
Published: January 8, 2017. **doi:** 10.12732/caa.v21i1.2

Dynamic Publishers, Inc., Acad. Publishers, Ltd. http://www.acadsol.eu/caa

1. INTRODUCTION

Consider the unconstrained optimization problem

$$\{\min f(x), \quad x \in \mathbb{R}^n\}, \tag{1.1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is continuously differentiable. The line search method usually takes the following iterative formula

$$x_{k+1} = x_k + \alpha_k d_k \tag{1.2}$$

for (1.1), where x_k is the current iterate point, $\alpha_k > 0$ is a steplength and d_k is a search direction. Different choices of d_k and α_k will determine different line search methods([24,26,27]). We denote $f(x_k)$ by f_k ,$\nabla f(x_k)$ by g_k, and $\nabla f(x_{k+1})$ by g_{k+1}, respectively. $\|.\|$ denotes the Euclidian norm of vectors and define $y_k = g_{k+1} - g_k$.

We all know that a method is called steepest descent method if we take $d_k = -g_k$ as a search direction at every iteration, which has wide applications in solving large-scale minimization problems ([24, 25,29]). One drawback of the method is often yielding zigzag phenomena in solving practical problems, which makes the algorithm converge to an optimal solution very slowly, or even fail to converge ([17, 19]).

If we take $d_k = -H_k g_k$ as a search direction at each iteration in the algorithm, where H_k is an $n \times n$ matrix approximating $\left[\nabla^2 f(x_k)\right]^{-1}$, then the corresponding method is called the Newton-like method ([17, 19, 29]) such as the Newton method, the quasi-Newton method, variable metric method, etc. Many papers have proposed this method for optimization problems ([5, 6, 9, 20]).

However, the Newton-like method needs to store and compute matrix H_k at each iteration and thus adds to the cost of storage and computation. Accordingly, this method is not suitable to solve large-scale optimization problems in many cases.

Due to its simplicity and its very low memory requirement, the conjugate gradient method is a powerful line search method for solving the large-scale optimization problems. In fact, the CG method is not among the fastest or most robust optimization algorithms for nonlinear problems available today, but it remains very popular for engineers and mathematicians who are interested in solving large problems (cf. [1,16,18,28]). The conjugate gradient method is designed to solve unconstrained optimization problem (1.1). More explicitly, the conjugate gradient method is an algorithm for finding the nearest local minimum of a function of variables which presupposes that the gradient of the function can be computed. We consider only the case where the method is implemented without regular restarts. The iterative formula of the conjugate gradient method is given by (1.2), where α_k is a steplength which is computed by carrying out a line search, and d_k is the search direction defined by

$$d_{k+1} = \begin{cases} -g_k, & \text{if } k = 1, \\ -g_{k+1} + \beta_k d_k, & \text{if } k \geq 2. \end{cases} \tag{1.3}$$

where β_k is a scalar and $g(x)$ denotes $\nabla f(x)$. If f is a strictly convex quadratic function, namely,

$$f(x) = \frac{1}{2} x^T H x + b^T x, \tag{1.4}$$

where H is a positive definite matrix and if α_k is the exact one-dimensional minimizer along the direction d_k, i.e.,

$$\alpha_k = \arg\min_{\alpha>0} \{f(x + \alpha d_k\} \tag{1.5}$$

then (1.2)–(1.3) is called the linear conjugate gradient method. Otherwise, (1.2)–(1.3) is called the nonlinear conjugate gradient method.

Conjugate gradient methods differ in their way of defining the scalar parameter β_k. In the literature, there have been proposed several choices for β_k which give rise to distinct conjugate gradient methods. The most well known conjugate gradient methods are the Hestenes–Stiefel (HS) method [12], the Fletcher–Reeves (FR) method [10], the Polak-Ribière-Polyak (PR) method [21,23], the Conjugate Descent method(CD) [9], the Liu-Storey (LS) method [15], the Dai-Yuan (DY) method [7], and Hager and Zhang (HZ) method [13]. The update parameters of these methods are respectively specified as follows:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k},\ \beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2},\ \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2},\ \beta_k^{CD} = -\frac{\|g_{k+1}\|^2}{d_k^T g_k},$$

$$\beta_k^{LS} = -\frac{g_{k+1}^T y_k}{d_k^T g_k},\ \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k},\ \beta_k^{HZ} = \left(y_k - 2d_k \frac{\|y_k\|^2}{d_k^T y_k} \right)^T \frac{g_{k+1}}{d_k^T y_k}.$$

The convergence behavior of the above formulas with some line search conditions has been studied by many authors for many years. The FR method with an exact line search was proved to globally convergent on general functions by Zoutendijk 30]. However, the PRP method and the HS method with the exact line search are not globally convergent, see Powell's counterexample [22]. In the already-existing convergence analysis and implementations of the conjugate gradient method, the strong Wolfe conditions, namely,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{1.6}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq -\sigma g_k^T d_k, \tag{1.7}$$

where $\delta \in \left]0, \frac{1}{2}\right[$ and $\sigma \in]\delta, 1[$ are often imposed on the line search (see e.g. [1,2,3,8,11]).

In addition, the sufficient descent condition:

$$g_k^T d_k \leq -c \left\| g_k \right\|^2. \tag{1.8}$$

has often been used in the literature to analyze the global convergence of conjugate gradient methods with inexact line searches. For instance, Al-Baali [1], Toouati-Ahmed and Storey [3], Hu and Storey [14], Gilbert and Nocedal [11] analyzed the global convergence of algorithms related to the Fletcher–Reeves method with the strong Wolfe line search. Their convergence analysis used the sufficient descent condition (1.8). As for the algorithms related to the PRP method, Gilbert and Nocedal [11] investigated wide choices of β_k that resulted in globally convergent methods. In order for the sufficient descent condition to hold, they modified the strong Wolfe line search to the two-stage line search, the first stage is to find a point using the strong Wolfe line search, and the second stage is when, at that point the sufficient descent condition does not hold, more line search iterations will proceed until a new point satisfying the sufficient descent condition is found. They hinted that the sufficient descent condition may be crucial for conjugate gradient methods.

The main aim of this note is to show that the descent property holds for all k and the global convergence is achieved for an inexact line search.

In this work, we try to accelerate the convergence of the gradient method by introducing a new direction d_k^{TMR} defined as follows:

$$d_k^{TMR} = \begin{cases} -\alpha g_k & \text{if } k = 1 \\ -\alpha g_k + \gamma \beta_k d_{k-1} & \text{if } k \geq 2 \end{cases} \tag{1.9}$$

This paper is organized as follows. In the next section, the New algorithms are stated and descent property is presented. The global convergence of the new methods are established in Section 3. Numerical results and a conclusion are presented in Section 4 and in Section 5, respectively.

2. CGTMR ALGORITHM

In this section, we give the specific form of the proposed new conjugate gradient method. As reported before our search directions d_k^{TMR} are defined as follows:

$$d_k^{TMR} = \begin{cases} -\alpha g_k & \text{if } k = 1 \\ -\alpha g_k + \gamma\beta_k d_{k-1} & \text{if } k \geq 2 \end{cases} \tag{2.1}$$

if $\begin{cases} \alpha = \frac{1}{\|g_k\|^2} \\ \gamma = \frac{\|g_k\|^2}{\|g_{k+1}\|^2} \end{cases}$

Recently, M. Belloufi and R. benzine[4] proposed **CGM** with:

$$d_{k+1}^{BB} = \begin{cases} -\frac{g_1}{\|g_1\|^2} & \text{if } k = 1 \\ -\frac{1}{\|g_{k+1}\|^2} g_{k+1} + d_k & \text{if } k \geq 2 \end{cases} \tag{2.2}$$

If $\beta_k = \beta_k^{FR}$. Then:

$$d_k^{TMR} = \begin{cases} -\alpha g_k & \text{if } k = 1 \\ -\alpha g_k + \gamma\beta_k^{FR} d_{k-1} & \text{if } k \geq 2 \end{cases} \tag{2.3}$$

$$d_k^{TMR} = \begin{cases} -\frac{g_k}{\|g_k\|^2} & \text{if } k = 1 \\ -\frac{1}{\|g_k\|^2} g_k + d_{k-1} & \text{if } k \geq 2 \end{cases} \tag{2.4}$$

2.1. CGTMR ALGORITHM

The algorithm is given as follows:

Algorithm 1. *Step 0: Given $x_1 \in \mathbb{R}^n$, set $d_1^{TMR} = -\frac{g_1}{\|g_1\|^2}$, $k := 1$.*
Step 1: If $\|g_k\| = 0$ then stop else go to Step 2.

Step 2: Set $x_{k+1} = x_k + \alpha_k d_k^{TMR}$ *where* d_k^{TMR} *is defined by (2.4), and* α_k *is*

defined by (1.3) and (1.4).

Step 3. Set $k := k + 1$ *and go to Step 1.*

The following theorem indicates that, if α_k is computed by the Wolfe line search (1.3) and (1.4) , then the search direction d_k^{TMR} satisfies the descent property.

Theorem 1. *If the steplength* α_k *is computed by the Wolfe line search (1.3) and (1.4) with* $\delta < \sigma < \frac{1}{2}$, *, then for the proposed conjugate gradient method, the inequality*

$$-\sum_{j=0}^{k-1} \sigma^j \leq g_k^T d_k \leq -2 + \sum_{j=0}^{k-1} \sigma^j \tag{2.5}$$

holds for all k, *and hence the descent property*

$$g_k^T d_k < 0, \forall k \tag{2.6}$$

holds, as long as $g_k \neq 0$.

Proof. The proof is by induction. For $k = 1$ Equations (2.5) and (2.6) is clearly satisfied.

Now we suppose that (2.5) and (2.6) hold for any $k \geq 1$.

It follows from the definition (2.4) of d_{k+1} that

$$g_{k+1}^T d_{k+1} = -1 + g_{k+1}^T d_k \tag{2.7}$$

and hence from (1.4) and (2.6) that

$$-1 + \sigma g_k^T d_k \leq g_{k+1}^T d_{k+1} \leq -1 - \sigma g_k^T d_k \tag{2.8}$$

Also, by induction assumption (2.7), we have

$$-\sum_{j=0}^{k} \sigma^j = -1 - \sigma \sum_{j=0}^{k-1} \sigma^j \leq g_{k+1}^T d_{k+1}$$

$$\leq -1 + \sigma \sum_{j=0}^{k-1} \sigma^j = -2 + \sum_{j=0}^{k} \sigma^j$$

Then, (2.7) holds for $k+1$.

Since

$$g_{k+1}^T d_{k+1} \le -2 + \sum_{j=0}^{k} \sigma^j \tag{2.9}$$

and

$$\sum_{j=0}^{k} \sigma^j < \sum_{j=0}^{\infty} \sigma^j = \frac{1}{1-\sigma} \tag{2.10}$$

where $\sigma \in \left]0, \frac{1}{2}\right]$, it follows from $1-\sigma > \frac{1}{2}$ that $-2 + \sum_{j=0}^{k} \sigma^j < 0$. Hence, from (2.9), we obtain $g_{k+1}^T d_{k+1} < 0$. We complete the proof by induction. □

3. GLOBAL CONVERGENCE

In order to establish the global convergence of the proposed method, we assume that the following assumption always holds, i.e. Assumption 3.1 :

Assumption 3.1. Let f be twice continuously differentiable, and the level set

$$L = \{x \in \mathbb{R}^n \mid f(x) \le f(x_1)\}$$

be bounded.

Theorem 2. *Suppose that x_1 is a starting point for which Assumption 3.1 holds. Consider the New method (1.2) and (2.1). If the steplength α_k is computed by the strong Wolfe line search (1.3) and (1.4) with $\delta < \sigma < \frac{1}{2}$, then the method is globally convergent, i.e.,*

$$\liminf_{k \longrightarrow \infty} \|g_k\| = 0 \tag{3.1}$$

Proof. It is shown in theorem 1 that the descent property (2.6) holds for $\sigma \in \left]0, \frac{1}{2}\right]$, so from (1.4), (2.5), and (2.10) it follows that

$$\left|g_k^T d_{k-1}\right| \le -\sigma g_{k-1}^T d_{k-1} \le \sigma \sum_{j=0}^{k-2} \sigma^j = \sum_{j=0}^{k-1} \sigma^j \le \frac{\sigma}{1-\sigma} \tag{3.2}$$

Thus from the definition of d_k and using (3.2) we deduce that

$$\begin{aligned}\|d_k\|^2 &= \frac{1}{\|g_k\|^2} - \frac{2}{\|g_k\|^2} g_k^T d_{k-1} + \|d_{k-1}\|^2 \\ &\leq \frac{1}{\|g_k\|^2} + \frac{2\sigma}{1-\sigma}\frac{1}{\|g_k\|^2} + \|d_{k-1}\|^2 \\ &= \left(\frac{1+\sigma}{1-\sigma}\right)\frac{1}{\|g_k\|^2} + \|d_{k-1}\|^2 \qquad (3.3)\end{aligned}$$

By applying this relation repeatedly, it follows that

$$\begin{aligned}\|d_k\|^2 &\leq \left(\frac{1+\sigma}{1-\sigma}\right)\sum_{j=2}^{k}\frac{1}{\|g_j\|^2} + \frac{1}{\|g_1\|^2} \\ &\leq \left(\frac{1+\sigma}{1-\sigma}\right)\sum_{j=1}^{k}\frac{1}{\|g_j\|^2} \qquad (3.4)\end{aligned}$$

where we used the facts that

$$\frac{1}{\|g_1\|^2} \leq \left(\frac{1+\sigma}{1-\sigma}\right)\frac{1}{\|g_1\|^2}$$

Now we prove (3.1) by contradiction. It assumes that (3.1) does not hold, then there exists a constant $\varepsilon > 0$ such that

$$\|g_k\| \geq \varepsilon > 0 \qquad (3.5)$$

holds for all k sufficiently large. Since g_k is bounded above on the level set L, it follows from (3.4) that

$$\|d_k\|^2 \leq c_1 k \qquad (3.6)$$

where c_1 is a positive constant. From (2.5) and (2.10), we have

$$\begin{aligned}\cos\theta_k = -\frac{g_k^T d_k}{\|g_k\|\,\|d_k\|} &\geq \left(2 - \sum_{j=0}^{k-1}\sigma^j\right)\frac{1}{\|g_k\|\,\|d_k\|} \\ &\geq \left(\frac{1-2\sigma}{1-\sigma}\right)\frac{1}{\|g_k\|\,\|d_k\|} \qquad (3.7)\end{aligned}$$

Since $\sigma < \frac{1}{2}$, substituting (3.6) and (3.5) into (3.7) gives

$$\sum_k \cos^2\theta_k \geq \left(\frac{1-2\sigma}{1-\sigma}\right)^2 \sum_k \frac{1}{\|g_k\|^2\,\|d_k\|^2} \geq c_2 \sum_k \frac{1}{k} \qquad (3.8)$$

where c_2 is a positive constant. Therefore, the series $\sum_k \cos^2\theta_k$ is divergent.

Let M be an upper bound of $\left\|\nabla^2 f(x)\right\|$ on the level set L, then

$$g_{k+1}^T d_k = \left(g_k + a_k \nabla^2 f(x)\right)^T d_k \le g_k^T d_k + M a_k \left\|d_k\right\|^2$$

Thus by using (1.4) we obtain

$$a_k \ge -\frac{(1-\sigma)}{M\left\|d_k\right\|^2} g_k^T d_k \tag{3.9}$$

Substituting a_k of (3.9) into (1.3) gives

$$\begin{aligned} f_{k+1} &\le f_k - \frac{(1-\sigma)\,\delta}{M}\left(\frac{g_k^T d_k}{\left\|d_k\right\|}\right)^2 \\ &= f_k - c_3\left\|g_k\right\|^2\cos^2\theta_k, \end{aligned}$$

where $c_3 = \frac{(1-\sigma)\delta}{M} > 0$. Since $f(x)$ is bounded below, $\sum_k \left\|g_k\right\|^2\cos^2\theta_k$ converges, which indicates that $\sum_k \cos^2\theta_k$ converges by use of (3.5). This fact contradicts (3.8). We complete the proof. □

4. NUMERICAL RESULTS AND DISCUSSIONS

In this section we report some numerical results obtained with an implementation of the $CGTMR$ algorithm. For our numerical tests, we used test functions and Fortran programs from ([01],[03]). Considering the same criterias as in ([02]), the code is written in Fortran and compiled with f 90 on a Workstation Intel Pentium 4 with 2 GHz. We selected a number of 105 unconstrained optimization test functions in generalized or extended form [18] (some from CUTE library [03]). For each test function we have taken twenty (20) numerical experiments with the number of variables increasing as $n = 2, 10, 30, 50, 70, 100,\ 300, 500, 700, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000,$ 10000.

The algorithm implements the Wolfe line search conditions (1.3) and (1.4), and the same stopping criterion $\left\|\nabla f(x_k)\right\| < 10^{-6}$. In all the algorithms we

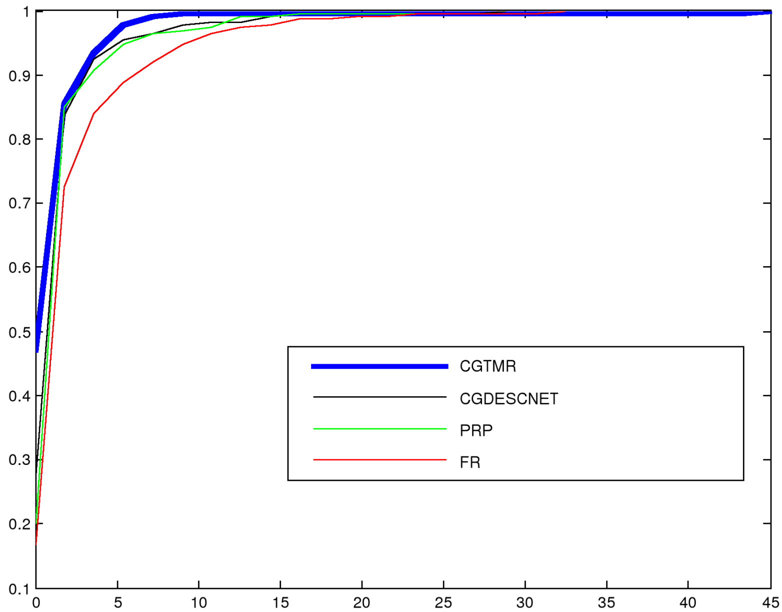

Figure 1: Performance based on CPU time

considered in this numerical study the maximum number of iterations is limited to 100000.

The comparisons of algorithms are given in the following context. Let f_i^{ALG1} and f_i^{ALG2} be the optimal value found by ALG1 and ALG2, for problem $i = 1, ..., 962$, respectively. We say that, in the particular problem i, the performance of ALG1 was better than the performance of ALG2 if:

$$\left| f_i^{ALG1} - f_i^{ALG2} \right| < 10^{-3}$$

and the number of iterations, or the number of function-gradient evaluations, or the CPU time of ALG1 was less than the number of iterations, or the number of function-gradient evaluations, or the CPU time corresponding to ALG2, respectively.

In a performance profile plot, the top curve corresponds to the method that solved the most problems in a time that was within a factor τ of the best time. The percentage of the test problems for which a method is the fastest is given on the left axis of the plot. The right side of the plot gives

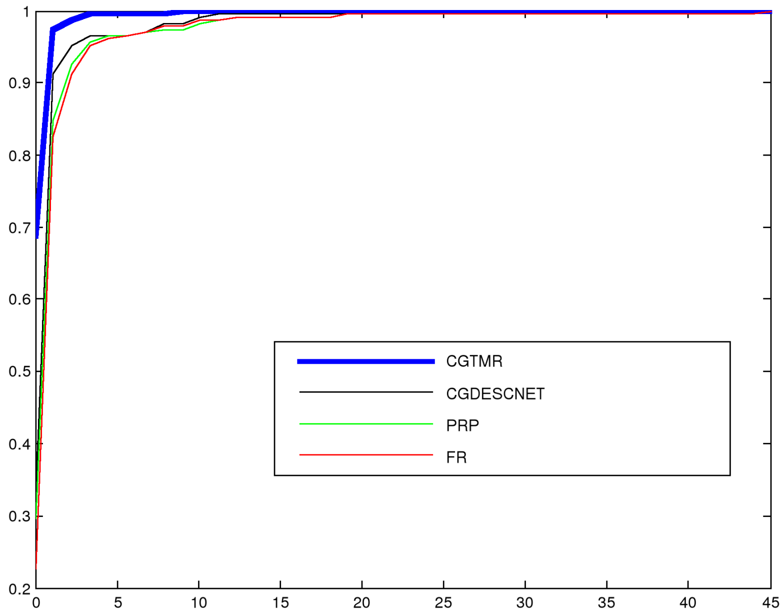

Figure 2: Performance based on the number of iterations

the percentage of the test problems that were successfully solved by these algorithms, respectively. Mainly, the right side is a measure of the robustness of an algorithm.

In the set of numerical experiments we compare $CGTMR$ algorithm to $CG\_DESCNET$, PRP and FR conjugate gradient methods.

Figs. $1-2$ list the performance of the $CGTMR$, $CG\_DESCNET$, PRP and FR conjugate gradient methods. relative to CPU time, the number of iterations, respectively, which were evaluated using the profiles of Dolan and Moré.

Clearly, Figs. $1-2$ present that our proposed method $CGTMR$ exhibits the best overall performance since it illustrates the highest probability of being the optimal solver, followed by the $CG\_DESCENT$, PRP and FR conjugate gradient methods relative to all performance metrics.

5. CONCLUSION

In this paper, however, we have proposed a new and simple $d_k$that is easy to implement. Our numerical results have shown that, our new method has the best performance compared to the other standard **CG** methods. We have also provided proof that this method converges globally with strong Wolfe line search.

REFERENCES

[1] M. Al-Baali, Descent property and global convergence of Fletcher-Reeves method with inexact line search, IMA. J. Numer. Anal. 5 (1985) 121-124.

[2] N. Andrei, An unconstrained optimization test functions collection, Adv. Modell. Optim. 10 (2008) 147-161.

[3] T. Ahmed, D. Storey, E cient hybrid conjugate gradient techniques, J. Optimiz. Theory Appl. 64 (1990) 379-394.

[4] M. Belloufi and R. benzine, Descent property andglobal convergence of a new search direction method for unconstrained optimization, Numerical Functional Analysis and Optimization,36:169–180, 2015, doi: 10.1080/01630563.2014.976796

[5] Y. Dai, Convergence properties of the BFGS algorithm. SIAM Journal on Optimization, 13 (2003) 693-701.

[6] J. E. Dennis, Jr., J. J. Mor e, (1977). Quasi-Newton methods, motivation and theory. SIAM Review, 19, 46-89.

[7] Y. Dai, Y. Yuan, A nonlinear conjugate gradient with a strong global convergence properties, SIAM J. Optimiz. 10 (2000) 177-182.

[8] Y. Dai, Y. Yuan, Convergence properties of the Fletch-Reeves method, MA J. Numer. Anal. 16 (2) (1996) 155-164.

[9] R. Fletcher, Practical Method of Optimization, second ed., Unconstrained Optimization, vol. I, Wiley, New York, 1997.

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

[11] J.C. Gibert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optimiz. 2 (1992) 21-42.

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

[13] W.W. Hager, H. Zhang, A new conjugate gradient method with guaranteed descent and an e cient line search, SIAM Journal on Optimization 16 (2005) 170-192.

[14] Y.F. Hu, C. Storey, Global convergence result for conjugate gradient method, J. Optimiz. Theory Appl. 71 (1991) 399-405.

[15] Y. Liu, C. Storey, E cient generalized conjugate gradient algorithms. Part 1: Theory, J. Optimiz. Theory Appl. 69 (1992) 129-137.

[16] G. Liu, J. Han, H. Yin, Global convergence of the Fletcher-Reeves algorithm with inexact line search, Appl. Math. JCU 10B (1995) 75-82.

[17] D. C. Luenerger, Linear and nonlinear programming (2nd ed.). Reading,MA: Addition Wesley, 1989.

[18] J. Nocedal, Conjugate gradient methods and nonlinear optimization, in: L. Adams, J.L. Nazareth (Eds.), Linear and Nonlinear Conjugate Gradient Related Methods, SIAM, Philadelphia, PA, 1995, 9-23.

[19] J. Nocedal, S.J. Wright, Numerical optimization, Springer Series in Operations Research, Springer, New York, 1999.

[20] M.J.D. Powell, A new algorithm for unconstrained optimization. In J. B. Rosen, O. L. Mangasarian, & K. Ritter (Eds.), Nonlinear programming. New York: Academic Press, (1970).

[21] B.T. Polyak, The conjugate gradient method in extreme problems, USSR Comp. Math. Math. Phys. 9 (1969) 94-112.

[22] M.J.D. Powell, Nonconvex minimization calculations and the conjugate gradient method, Lecture Notes in Mathematics, 1066, Springer-Verlag, Berlin, 1984, 122-141.

[23] E. Polak, G. Ribi ere, Note Sur la convergence de directions conjug ees, Rev. Francaise Informat Recherche Operationelle 3e Ann ee 16(1969) 35-43.

[24] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM Journal of Optimization, 7 (1997) 26-33.

[25] M. Raydan, B. F. Svaiter, Relaxed steepest descent and Cauchy-Barzilai-Borwein method. Computational Optimization and Applications, 21 (2002) 155-167.

[26] J. Schropp, A note on minimization problems and multistep methods. Numeric Mathematic, 78 (1997) 87-101.

[27] J. Schropp, One-step and multistep procedures for constrained minimization problems. IMA Journal of Numerical Analysis, 20 (2000) 135-152.

[28] J. Sun, J. Zhang, Convergence of conjugate gradient methods without line search, Annals of Operations Research 103 (2001) 161-173.

[29] Y. Yuan, W. Sun, Theory and Methods of Optimization, Science Press ofChina, Beijing, 1999.

[30] G. Zoutendijk, Nonlinear programming computational methods, in: J. Abadie (Ed.), Integer and Nonlinear Programming, North-Holland, Amsterdam, 1970, 37-86.