Convergence of Three-term Conjugate Gradient Methods*

Y. H. Dai and Y. Yuan

State Key Laboratory of Scientific and Engineering Computing,
Institute of Computational Mathematics and Scientific/Engineering Computing,
Chinese Academy of Sciences, P. O. Box 2719, Beijing 100080, China
Email-addresses: dyh@lsec.cc.ac.cn, yyx@lsec.cc.ac.cn

Abstract

This paper studies the three-term conjugate gradient method for unconstrained optimization. The method includes the classical (two-term) conjugate gradient method and the famous Beale-Powell restart algorithm as its special forms. Some mild conditions are given in this paper, which ensure the global convergence of general three-term conjugate gradient methods.

Key words: unconstrained optimization, conjugate gradient, line search, global convergence.

AMS classification: 65k, 90c.

1. Introduction

To solve the following optimization problem

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

we consider the iterative method

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

where x_1 is given, d_k is a search direction, and α_k is a steplength. In the classical conjugate gradient method, the search direction d_k ($k \geq 2$) is defined by the current negative gradient $-g_k$ and the previous search direction d_{k-1}, namely,

$$d_k = -g_k + \beta_k d_{k-1}, \tag{1.3}$$

where $d_1 = -g_1$, and β_k is a scalar. Two famous formulae for β_k are called the Fletcher-Reeves (FR), and the Polak-Ribiére-Polyak (PRP) formulae (see [6; 9, 10]), and they are given by

$$\beta_k^{FR} = \|g_k\|^2 / \|g_{k-1}\|^2 \tag{1.4}$$

*Research partially supported by Chinese NSF grants 19525101, 19731010 and 19801033

and

$$\beta_k^{PRP} = g_k^T(g_k - g_{k-1})/\|g_{k-1}\|^2, \tag{1.5}$$

respectively. Here $\|\cdot\|$ stands for the two norm. In [1], Beale proposed a three-term conjugate gradient method, in which the search direction d_k has the form

$$d_k = -g_k + \beta_k d_{k-1} + \gamma_k d_t, \tag{1.6}$$

where d_t is a restart direction. Powell[11] established efficient restart strategy for this method and obtained good numerical results. In this paper, we will study the following general three-term conjugate gradient method:

$$d_k = -g_k + \beta_k d_{k-1} + \gamma_k d_{t(p)}, \tag{1.7}$$

where $t(p)$ is the number of the p-th restart iteration satisfying $t(p) < k \le t(p+1) \le +\infty$. It is obvious that the method (1.7) includes the classical (two-term) conjugate gradient method and the Beale-Powell restart algorithm as its special forms.

References [3] and [2] have analyzed the global convergence properties of general two-term conjugate gradient methods. [5] extended [8]'s convergence result on the FR method to general three-term conjugate gradient methods. [4] pointed out that the Beale-Powell restart algorithm needs not converge for general objective functions. However, with slight modifications, the Beale-Powell restart algorithm is proved to be globally convergent. In this paper, we will give some mild conditions which ensure the global convergence of general three-term conjugate gradient methods.

2. Preliminaries

Throughout this paper, we assume that

$$g_k \neq 0, \quad \text{for all } k \ge 1, \tag{2.1}$$

for otherwise a stationary point has been found. We also assume that the objective function satisfies the following assumption.

Assumption 2.1 *(i) The level set $\mathcal{L} = \{x \in \Re^n : f(x) \le f(x_1)\}$ is bounded; (ii) In some neighborhood $\mathcal{N}$ of $\mathcal{L}$, f is differentiable and its gradient g is Lipschitz continuous, namely, there exists a constant $L > 0$ such that*

$$\|g(x) - g(\tilde{x})\| \le L\|x - \tilde{x}\|, \qquad \textit{for all } x,\ \tilde{x} \in \mathcal{N}. \tag{2.2}$$

The steplength α_k in (1.2) is often computed by mean of a one-dimensional line search. The standard Wolfe line search ([14]) is to compute α_k such that

$$\begin{aligned} f(x_k) - f(x_k + \alpha_k d_k) &\ \ge\ -\delta\alpha_k g_k^T d_k, && (2.3)\\ g(x_k + \alpha_k d_k)^T d_k &\ \ge\ \sigma g_k^T d_k, && (2.4) \end{aligned}$$

where $0 < \delta < \sigma < 1$. For the standard Wolfe line search, we state the following lemma, which was essentially proved by Wolfe[14, 15] and Zoutendijk[16].

Lemma 2.2 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the iterative method (1.2), where d_k is a descent direction and α_k satisfies (2.3)–(2.4). Then we have that*

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{2.5}$$

In many implementations and theoretical analyses of conjugate gradient methods, the steplength α_k is often obtained by the strong Wolfe line search. The strong Wolfe line search is to compute α_k satisfying (2.3) and

$$|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k, \tag{2.6}$$

where also $0 < \delta < \sigma < 1$. Relation (2.6) is stronger than (2.4), which implies that Lemma 2.2 also holds for the strong Wolfe line search. In this paper, we will study the global convergence of the method (1.7) using strong Wolfe line searches.

In the analyses in the coming sections, we still need the following lemma. One can see [13] for its proof.

Lemma 2.3 *Suppose that $\{a_i\}$ and $\{b_i\}$ are two positive sequences. If*

$$\sum_{k\geq 1} a_k = \infty, \tag{2.7}$$

and there exist two constants c_1 and c_2 such that for all $k \geq 1$,

$$b_k \leq c_1 + c_2 \sum_{i=1}^{k} a_i, \tag{2.8}$$

then we have that

$$\sum_{k\geq 1} \frac{a_k}{b_k} = \infty. \tag{2.9}$$

3. Convergence of the method (1.7)

In the following, we will first analyze the general three-term conjugate gradient method using strong Wolfe line searches.

Theorem 3.1 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the general three-term conjugate gradient method (1.2) and (1.7), where the steplength α_k is such that (2.3), (2.6), and the descent condition $g_{k+1}^T d_{k+1} < 0$ hold. Assume that there exist two constants σ_1, $\sigma_2 > 0$ such that the following relations hold*

$$\|\gamma_k d_{t(p)}\| \leq \sigma_1 \|g_k\|, \tag{3.1}$$

$$|\gamma_k g_k^T d_{t(p)}| \leq \sigma_2 |\beta_k g_k^T d_{k-1}|. \tag{3.2}$$

Then if

$$\sum_{k\geq 1} \frac{\|g_k\|^t}{\|d_k\|^2} = \infty \tag{3.3}$$

holds for some constant $t \in [0, 4]$, the method converges in the sense that

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.4}$$

Proof ¿From (1.7), the triangle inequality and (3.1), we have that

$$\begin{aligned} \|d_k\| &\geq \|\beta_k d_{k-1}\| - \| - g_k + \gamma_k d_{t(p)}\| \\ &\geq |\beta_k| \|d_{k-1}\| - (1+\sigma_1)\|g_k\|. \end{aligned} \tag{3.5}$$

Multiplying (3.5) with g_k and using (3.2) and (2.6), we get that

$$\begin{aligned} \|g_k\|^2 &= -g_k^T d_k + \beta_k g_k^T d_{k-1} + \gamma_k g_k^T d_{t(p)} \\ &\leq |g_k^T d_k| + (1+\sigma_2)|\beta_k g_k^T d_{k-1}| \\ &\leq |g_k^T d_k| + c|\beta_k g_{k-1}^T d_{k-1}|, \end{aligned} \tag{3.6}$$

where $c = \sigma(1+\sigma_2)$. Define

$$\tau_k = \frac{|g_k^T d_k|}{\|d_k\|}. \tag{3.7}$$

It follows by (3.6) and (3.5) that

$$\begin{aligned} \frac{\|g_k\|^2}{\|d_k\|} &\leq \tau_k + c\tau_{k-1} \frac{|\beta_k| \|d_{k-1}\|}{\|d_k\|} \\ &\leq \tau_k + c\tau_{k-1} \left[1 + (1+\sigma_1) \frac{\|g_k\|}{\|d_k\|} \right]. \end{aligned} \tag{3.8}$$

Thus we have that

$$\frac{\|g_k\|^2}{\|d_k\|} \left[1 - c(1+\sigma_1)\|g_k\|^{-1} \tau_{k-1} \right] \leq \tau_k + c\tau_{k-1}. \tag{3.9}$$

On the other hand, by the definition (3.7) of τ_k, (2.5) can be rewritten as

$$\sum_{k \geq 1} \tau_k^2 < \infty, \tag{3.10}$$

which shows that

$$\lim_{k \to \infty} \tau_k = 0. \tag{3.11}$$

Now we suppose that (3.4) does not hold and there exists a constant $\gamma > 0$ such that

$$\|g_k\| \geq \gamma, \quad \text{for all } k \geq 1. \tag{3.12}$$

Then by this and (3.11), there must exist some constant k_0 such that

$$c(1+\sigma_1)\tau_{k-1} \leq \frac{1}{2}\|g_k\| \tag{3.13}$$

holds for all $k \geq k_0$. It follows from (3.9), (3.13) and (3.10) that

$$\sum_{k \geq k_0+1} \frac{\|g_k\|^4}{\|d_k\|^2} \leq 4 \sum_{k \geq k_0+1} (\tau_k + c\tau_{k-1})^2 \leq 8(1+c^2) \sum_{k \geq k_0} \tau_k^2 < \infty. \tag{3.14}$$

The above relation contradicts (3.3), since $t \in [0, 4]$ and by (3.12), $\|g_k\|$ is bounded away from zero. The contradiction shows the truth of (3.4). $\square$

It is easy to see that the assumptions used in the above theorem are weaker than those in [5]. If the constant σ_1 in (3.1) satisfies

$$\sigma_1 < 1, \tag{3.15}$$

then instead of (3.2), we can show by (1.7), (2.6) and (3.15) that

$$(1-\sigma_1)\|g_k\|^2 \le |g_k^T d_k| + \sigma|\beta_k g_{k-1}^T d_{k-1}|. \tag{3.16}$$

Thus Theorem 3.1 is also true if condition (3.2) is replaced by (3.15).

Theorem 3.2 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the general three-term conjugate gradient method (1.2) and (1.7), where the steplength α_k is such that (2.3), (2.6), and the descent condition $g_{k+1}^T d_{k+1} < 0$ hold. If relations (3.1), (3.3) and (3.15) hold, where $t \in [0,4]$, the method converges in the sense that (3.4) is true.*

By Theorem 3.1, we can prove the following convergence result.

Corollary 3.3 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the general three-term conjugate gradient method (1.2) and (1.7), where the steplength α_k is such that (2.3), (2.6), and the descent condition $g_{k+1}^T d_{k+1} < 0$ hold. If relations (3.1), (3.15) and*

$$\sum_{k\ge 1} \frac{|g_k^T d_k|^r}{\|d_k\|^2} = \infty \tag{3.17}$$

hold, where $r \in [0,2]$ is constant, the method converges in the sense that (3.4) is true.

Proof For any $r \in [0,2]$, if $|g_k^T d_k| > 1$, then $|g_k^T d_k|^r \le (g_k^T d_k)^2$. Thus we always have that

$$|g_k^T d_k|^r \le 1 + (g_k^T d_k)^2, \tag{3.18}$$

which implies that

$$\sum_{k\ge 1} \frac{1}{\|d_k\|^2} \ge \sum_{k\ge 1} \frac{|g_k^T d_k|^r}{\|d_k\|^2} - \sum_{k\ge 1} \frac{|g_k^T d_k|^2}{\|d_k\|^2}. \tag{3.19}$$

The above relation, Lemma 2.2 and (3.17) show that (3.3) holds with $t = 0$. Therefore by Theorem 3.1, (3.4) is true. □

Using Theorem 3.1 and Corollary 3.3, we now are able to give a sufficient condition that ensures the global convergence of the general three-term conjugate gradient method using strong Wolfe line searches. The condition concerns about the scalar β_k only.

Theorem 3.4 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the general three-term conjugate gradient method (1.2) and (1.7), where the steplength α_k is such that (2.3), (2.6), and the descent condition $g_{k+1}^T d_{k+1} < 0$ hold. Assume that (3.1), (3.2) and*

$$\beta_k \gamma_k d_k^T d_{t(p)} \le 0 \tag{3.20}$$

holds for all $p \ge 1$ and $t(p) < k \le t(p+1)$. If there exist an infinite sequence $\{k_i\} \subset \{1,2,3,\ldots\}$ and a constant $c_1 > 0$ such that

$$\prod_{j=l}^{k_i} |\beta_j| \le c_1 \tag{3.21}$$

holds for all $i \ge 1$ and $l \le k_i$, the method converges in the sense that (3.4) is true.

Proof We rewrite (1.7) as

$$d_k + g_k = \beta_k d_{k-1} + \gamma_k d_{t(p)}. \tag{3.22}$$

Taking norms and using (3.1) and (3.20), we can get

$$\|d_k\|^2 \le (\sigma_1^2 - 1)\|g_k\|^2 - 2g_k^T d_k + \beta_k^2 \|d_{k-1}\|^2. \tag{3.23}$$

The recursion of (3.23) yields

$$\|d_k\|^2 \le (\sigma_1^2 - 1)\|g_k\|^2 - 2g_k^T d_k + \sum_{j=2}^{k} \prod_{i=j}^{k} \beta_i^2 [(\sigma_1^2 - 1)\|g_j\|^2 - 2g_j^T d_j], \tag{3.24}$$

which with (3.21) implies that

$$\|d_{k_i}\|^2 \le (1 + c_1^2) \sum_{i=1}^{k_i} [(\sigma_1^2 - 1)\|g_i\|^2 - 2g_i^T d_i]. \tag{3.25}$$

If

$$\liminf_{i\to\infty} \|d_{k_i}\| < \infty, \tag{3.26}$$

then (3.3) holds with $t = 0$ and hence (3.4) follows Theorem 3.1. Thus we can assume

$$\lim_{i\to\infty} \|d_{k_i}\| = \infty. \tag{3.27}$$

In this case, we can prove by (3.25) and Lemma 2.3 that

$$\lim_{i\to\infty} \sum_{j=1}^{k_i} \frac{(\sigma_1^2 - 1)\|g_j\|^2 - 2g_j^T d_j}{\|d_j\|^2} = \infty. \tag{3.28}$$

Therefore either (3.3) holds with $t = 2$, or (3.17) holds with $r = 1$. By Theorem 3.1 and Corollary 3.3, we must have (3.4). □

In the above results, we do not make any assumptions on the restarts of the three-term conjugate gradient method. In the following, we will consider the method under certain restart criterions. Firstly, because of the quadratic termination of the conjugate gradient method, we assume that the method will be restarted at most every n steps, where n is the dimension of the objection function. In other words, we assume

$$t(p+1) - t(p) \le n. \tag{3.29}$$

Secondly, in the case when the previous search direction d_{k-1} tends to be opposite to the restart direction $d_{t(p)}$, it is reasonable to discard the second derivative information that is found along $d_{t(p)}$. More exactly, we also restart the method if the following relation does not hold,

$$\frac{d_{k-1}^T d_{t(p)}}{\|d_{k-1}\|\|d_{t(p)}\|} \ge -c_2, \tag{3.30}$$

where $c_2 \in (0, 1)$ is constant. In addition, if the following relation is not true,

$$|g_k^T d_{t(p)}| \le \sigma_3 |g_{t(p)}^T d_{t(p)}|, \tag{3.31}$$

where σ_3 is some positive constant, we also restart the method. For the restart criterion given as above, we can prove the following theorem.

Theorem 3.5 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the general three-term conjugate gradient method (1.2) and (1.7), where the steplength α_k is such that (2.3), (2.6), and the descent condition $g_{k+1}^T d_{k+1} < 0$ hold. Assume that $t(p)$ satisfies (3.29), and the method is restarted if either of the relations (3.30) and (3.31) does not hold. If*

$$\sum_{k\geq 1} \frac{\|g_k\|^t}{\|d_k\|^2} = \infty \tag{3.32}$$

holds for some constant $t \in [0, 4]$, the method converges in the sense that (3.4) is true.

Proof Denote N to be the set of all positive integers and define

$$N_1 = \{k \in N : \; g_k^T d_k \leq -\frac{1}{2}\|g_k\|^2\}, \tag{3.33}$$

$$N_2 = \{k \in N \backslash N_1 : \; k = t(p) + 1 \text{ for some } p\} \tag{3.34}$$

and

$$N_3 = N \backslash (N_1 \bigcup N_2). \tag{3.35}$$

For any integer $k \in N_1$, we clearly have from Lemma 2.2 that

$$\sum_{k\in N_1} \frac{\|g_k\|^4}{\|d_k\|^2} \leq 4 \sum_{k\in N_1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{3.36}$$

In the case when $k \in N_2$, since d_k is defined only by $-g_k$ and d_{k-1}, we can prove similarly to the proof of Theorem 2.3 in [3] that, if (3.12) holds, there exist constant $c_3 > 0$ and integer k_0 such that

$$\frac{\|g_k\|^4}{\|d_k\|^2} \leq \frac{2}{c_3} \left[\frac{(g_k^T d_k)^2}{\|d_k\|^2} + \frac{(g_{k-1}^T d_{k-1})^2}{\|d_{k-1}\|^2} \right] \tag{3.37}$$

holds for all $k \geq k_0$. Thus by Lemma 2.2, we also have that

$$\sum_{k\in N_2} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{3.38}$$

Now we assume that $k \in N_3$. In this case, by the definition of N_3 and the assumptions in this theorem, we have that (3.30), (3.31) and

$$g_k^T d_k > -\frac{1}{2}\|g_k\|^2 \tag{3.39}$$

hold for $k \in N_3$. Taking norms in (3.22), we get that

$$\|d_k\|^2 = -\|g_k\|^2 - 2g_k^T d_k + \beta_k^2 \|d_{k-1}\|^2 + 2\beta_k \gamma_k d_{k-1}^T d_{t(p)} + \gamma_k^2 \|d_{t(p)}\|^2. \tag{3.40}$$

By this and (3.39), it is easy to show

$$\|d_k\|^2 \leq 2(\beta_k^2 \|d_{k-1}\|^2 + \gamma_k^2 \|d_{t(p)}\|^2). \tag{3.41}$$

On the other hand, it follows from (1.7), (2.6) and (3.31) that

$$\|g_k\|^2 \leq |g_k^T d_k| + \sigma |\beta_k g_{k-1}^T d_{k-1}| + \sigma_3 |\gamma_k g_{t(p)}^T d_{t(p)}|, \tag{3.42}$$

which implies that

$$\|g_k\|^4 \le 2[(g_k^T d_k)^2 + \sigma^2\beta_k^2(g_{k-1}^T d_{k-1})^2 + \sigma_3^2\gamma_k^2(g_{t(p)}^T d_{t(p)})^2]. \tag{3.43}$$

Combining (3.41) and (3.43), we obtain

$$\begin{aligned}\frac{2(g_k^T d_k)^2}{\|d_k\|^2} &+ \frac{\sigma^2(g_{k-1}^T d_{k-1})^2}{\|d_{k-1}\|^2} + \frac{\sigma_3^2(g_{t(p)}^T d_{t(p)})^2}{\|d_{t(p)}\|^2} \\ &\ge \frac{(g_k^T d_k)^2 + \sigma^2\beta_k^2(g_{k-1}^T d_{k-1})^2 + \sigma_3^2\gamma_k^2(g_{t(p)}^T d_{t(p)})^2}{\beta_k^2\|d_{k-1}\|^2 + \gamma_k^2\|d_{t(p)}\|^2} \\ &\ge \frac{\|g_k\|^4}{2\beta_k^2\|d_{k-1}\|^2 + 2\gamma_k^2\|d_{t(p)}\|^2}. \end{aligned} \tag{3.44}$$

By the above relation, (3.29) and Lemma 2.2, we have

$$\sum_{k\in N_3} \frac{\|g_k\|^4}{\beta_k^2\|d_{k-1}\|^2 + \gamma_k^2\|d_{t(p)}\|^2} \le c_4 \sum_{k\ge 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty, \tag{3.45}$$

where $c_4 = 2[2 + \sigma^2 + n\sigma_3^2]$. Thus if (3.12) holds, we know from (3.45) that for all large k,

$$\beta_k^2\|d_{k-1}\|^2 + \gamma_k^2\|d_{t(p)}\|^2 \ge \frac{2\|g_k\|^2}{1-\delta}. \tag{3.46}$$

This and (3.30) show that

$$\begin{aligned}\|d_k\|^2 &\ge -\|g_k\|^2 + (1-c_2)(\beta_k^2\|d_{k-1}\|^2 + \gamma_k^2\|d_{t(p)}\|^2) \\ &\ge \frac{1-c_2}{2}(\beta_k^2\|d_{k-1}\|^2 + \gamma_k^2\|d_{t(p)}\|^2), \end{aligned} \tag{3.47}$$

which with (3.45) indicates

$$\sum_{k\in N_3} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{3.48}$$

Therefore, if (3.12) holds, we have from (3.36), (3.38) and (3.48) that

$$\sum_{k\ge 1} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{3.49}$$

The above relation and (3.32) give a contradiction. Thus (3.4) must hold. □

4. Some discussions

We have proposed some mild conditions which ensure the global convergence of the general three-term conjugate gradient method using strong Wolfe line searches. These conditions are weaker than those used in reference [5]. The results of this paper provide some unified approaches to the analyses of the three-term conjugate gradient method, since they do not make any assumptions on the restarts.

Since the three-term conjugate gradient method includes the classical two-term conjugate gradient method as its special form, some results made in [3] and [2] can be also regarded corollaries of this paper. In addition, according to the discussions in the last section of [2], one can see that condition (3.21) can not be relaxed.

Some of our attentions were also devoted to the restart criterions in the three-term conjugate gradient method. It is expected that besides (3.29), the other criterions (3.30) and (3.31) will be also helpful in designing more efficient three-term conjugate gradient methods.

References

[1] Beale E M L. A derivative of conjugate gradients. in: Lootsma F A, eds. Numerical methods for nonlinear optimization, London, Academic Press, 1972, 39-43

[2] Dai Y H. and Yuan Y X. A Note on the nonlinear conjugate gradient method. Research Report ICM 98-012, Inst. of Comp. Math. and Sci. Eng., Chinese Academy of Sciences, 1998.

[3] Dai Y H, Han J Y, Liu G H, Sun D F, Yin H X, Yuan Y. Convergence properties of nonlinear conjugate gradient methods. 1998. (Accepted by SIAM Journal of Optimization)

[4] Dai Y H, Yuan Y. Convergence properties of the Beale-Powell restart algorithm. Science in China (Series A), 1998, 41(11): 1142-1150.

[5] Deng N, Li Z. Global convergence of three terms conjugate gradient methods. Optimization Methods and Sofeware, 1995, 4: 273-282

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

[7] Gilbert J C, Nocedal J. Global convergence properties of conjugate gradient methods for optimization. SIAM J Optimization, 1992, 2(1): 21-42

[8] Hu Y F, Storey C. Global convergence result for conjugate gradient methods, J. Optim. Theory Appl. 1991, 49: 409-436

[9] Polak E, Ribière G. Note sur la convergence de directions conjugées. Rev Francaise Informat Recherche Opertionelle, 1969, 3(16): 35-43

[10] Polyak B T. The conjugate gradient method in extremem problems. USSR Comp Math and Math Phys, 1969, 9: 94-112

[11] Powell M J D. Restart procedures of the conjugate gradient method. Math Program, 1977, 2: 241-254

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

[13] Pu D and Yu W, On the convergence properties of the DFP algorithms, Annals of Operations Research 1990, 24: 175-184.

[14] Wolfe P. Convergence conditions for ascent methods. SIAM Review, 1969, 11: 226-235

[15] Wolfe P. Convergence conditions for ascent methods. II: Some corrections. SIAM Review, 1969, 13: 185-188

[16] Zoutendijk G. Nonlinear Programming, Computational Methods. in: Abadie J, eds. Integer and Nonlinear Programming, North-Holland, Amsterdam, 1970, 37-86