Journal of Computational Mathematics, Vol.19, No.5, 2001, 539–548.

CONVERGENCE OF NONLINEAR CONJUGATE GRADIENT METHODS*[1)]

Yu-hong Dai
(*LSEC, ICMSEC, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing 100080, China*)

Abstract

This paper proves that a simplified Armijo-type line search can ensure the global convergences of the Fletcher-Reeves method and the Polak-Ribiére-Polyak method for unconstrained optimization. Although it seems not possible to verify that the PRP method using the generalized Armijo line search converges globally for generally problems, it can be shown that in this case the PRP method always solves uniformly convex problems.

Key words: Unconstrained optimization, Conjugate gradient, (generalized) Line search, Global convergence.

1. Introduction

Consider the unconstrained optimization problem,

$$\min \ f(x), \tag{1.1}$$

where f is smooth and its gradient g is available. Conjugate gradient methods are highly useful for solving (1.1) especially if n is large. They are iterative methods of the form

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

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

Here α_k is a stepsize obtained by a 1-dimensional line search and β_k is a scalar. The choice of β_k is such that (1.2)–(1.3) reduces to the linear conjugate gradient method in the case when f is a strictly convex quadratic and α_k is the exact 1-dimensional minimizer. The first nonlinear conjugate gradient method is presented by Fletcher and Reeves [11] in 1964, and has the following formula for β_k:

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

where and below we use $\|\cdot\|$ for the two norm. Another well-known formula for β_k is

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

which is proposed by Polak and Ribiére [22] and Polyak [23] in 1969 independently. For simplicity, we call the methods (1.2)–(1.3) where β_k are given by (1.4) and (1.5) as the FR method and the PRP method respectively. See [6, 9, 10, 15, 18] for some other choices for β_k. Nice reviews of the nonlinear conjugate gradient method can be seen in [20] and [21]. In this paper, our attention will be paid to the FR method and the PRP method only.

* Received December 29, 1997.
[1)] This work was supported by the Chinese NSF grant 19801033 and a youth innovation fund of CAS.

The FR method has been studied in many references, including [1, 5, 7, 8, 12, 19, 24, 27, 28, 30]. It is generally believed that the FR method has nice global convergence properties though it performs often much slower than the PRP method. Recent results in [8] and [28] show that, if the objective function satisfies Assumption 2.1 and has bounded level sets, and if each search direction is a descent direction, then the FR method using the standard Wolfe line search or the standard Armijo line search converges in the sense that

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{1.6}$$

(see [29] and [2] for the two line searches.) As compared with the FR method, despite its good numerical performances, the PRP method needs not converge to any stationary point even if the line search is exact (see [24]). In [12], Gilbert and Nocedal considered a suggestion in [25] of setting

$$\beta_k = \{\beta_k^{PRP}, 0\}, \tag{1.7}$$

and proved that such a modification results in (1.6). However, since as pointed out in [12], the value of β_k^{PRP} can be negative even in the case of strongly convex functions and exact line searches, Grippo and Lucidi [14] designed an Armijo-type line search for the PRP method, and showed that under some mild assumptions on f, the PRP method using the line search converges in the sense that

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{1.8}$$

The line search of Grippo and Lucidi is somewhat restrictive and complicated (see Algorithm 1 in [14]). Starting with an initial stepsize in the interval $[\rho_1\Delta_k,\ \rho_2\Delta_k]$, where $0 < \rho_1 < \rho_2$ and $\Delta_k = |g_k^T d_k|/\|d_k\|^2$, their line search multiplies the old trial stepsize by a constant in $(0,1)$ until the vectors $x_{k+1} = x_k + \alpha_k d_k$ and $d_{k+1} = -g_{k+1} + \beta_k^{PRP} d_k$ satisfy the following two condtions:

$$f_{k+1} \le f_k - \gamma \alpha_k^2 \|d_k\|^2 \tag{1.9}$$

and

$$-\delta_2 \|g_{k+1}\|^2 \le g_{k+1}^T d_{k+1} \le -\delta_1 \|g_{k+1}\|^2, \tag{1.10}$$

where $\gamma > 0$, $0 < \delta_1 < 1$ and $\delta_2 > 1$. Condition (1.9) is the basis of the line search techniques proposed in [17] and [13], in connection with no-derivative methods for unconstrainded optimization. Since one would usually be satisfied with any stationary point in real computations, in which case (1.6) and (1.8) can be regarded as the same, we wonder whether the line search of Grippo and Lucidi could be relaxed or not while only preserving (1.6) for the PRP method. Another motivation of this paper is that, since the FR method is generally believed to have better global convergence properties than the PRP method, we doubt if the FR method converges globally in the same case.

For the above reasons, we will study the FR method and the PRP method under a simplified Armijo-type line search. For the purpose of theoretical analyses, the generalized line search technique in [4] will be used in this paper to deal with the case when a descent search direction is not produced (see Section 2). From Theorems 3.3 and 3.4, one can see that the convergence properties of the FR method and the PRP method under the simplified Armijo-type line search are very satisfactory. Although it seems not possible to prove the convergence of the PRP method using the generalized Armijo line search for generally problems, we are able to show that in this case the PRP method converges to the unique minimizer if the objective function is uniformly convex. Some discussion is made in the last section.

2. Preliminaries

Assume that the objective function satisfies the following assumption.

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

$$\|g(x) - g(\tilde{x})\| \le L\|x - \tilde{x}\|, \qquad \text{for any } x,\ \tilde{x} \in \mathcal{N}. \tag{2.1}$$

Supposing that d_k is the k-th direction produced by any method in the form (1.2)–(1.3), we consider a generalized line search, as described as follows.

Line Search (A). *Given $\lambda \in (0,1)$ and $\delta > 0$. Determine the smallest integer $m \ge 0$ such that, if one defines*

$$\alpha_k = sign(-g_k^T d_k)\, \lambda^m, \tag{2.2}$$

then

$$f(x_k + \alpha_k d_k) - f_k \le -\delta \alpha_k^2 \|d_k\|^2. \tag{2.3}$$

If $g_k^T d_k < 0$, the above line search is one of the line search proposed in [17] and [13]. If $g_k^T d_k > 0$, line search (A) takes a negative stepsize along d_k. The third case is that d_k is orthogonal to g_k, in which $\alpha_k = 0$ is accepted and as a result, the next direction d_{k+1} produced by any method in the form (1.2)-(1.3) is a descent direction. Under Assumption 2.1 on f, we have the following lemma for line search (A).

Lemma 2.2. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider any method in the form (1.2), where d_k is an arbitrary direction and α_k is obtained by line search (A). Then if denoting $s_k = x_{k+1} - x_k$, there exists some constant $M > 0$ such that*

$$\sum_{k \ge 1} \|s_k\|^2 \le M. \tag{2.4}$$

Further, if denoting $N_1(k)$ and $N_2(k)$ as the sets of all integers i from 1 to k for which $|\alpha_i| = 1$ and $|\alpha_i| \in [0,1)$ respectively, we have that

$$\lim_{k\to\infty} \sum_{i \in N_1(k)} \|d_k\|^2 < \infty \tag{2.5}$$

and

$$\lim_{k\to\infty} \sum_{i \in N_2(k)} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{2.6}$$

Proof. Summing (2.3) for all k, we get that

$$\delta \sum_{k \ge 1} \|s_k\|^2 \le f_1 - f_{\min}, \tag{2.7}$$

where

$$f_{\min} = \min_{x \in \mathcal{L}} f(x). \tag{2.8}$$

Thus (2.4) holds with $M = (f_1 - f_{\min})/\delta$. For every k, if $|\alpha_k| \in (0,1)$, the line search implies that

$$f(x_k + \alpha_k d_k) - f_k \le -\delta \alpha_k^2 \|d_k\|^2 \tag{2.9}$$

and

$$f(x_k + \lambda^{-1} \alpha_k d_k) - f_k \ge -\delta \lambda^{-2} \alpha_k^2 \|d_k\|^2. \tag{2.10}$$

It follows from (2.10) that

$$f(x_k + \theta\alpha_k d_k) - f_k = -\delta\theta^2\alpha_k^2\|d_k\|^2 \tag{2.11}$$

for some $\theta \in [1, \lambda^{-1}]$. Define $\tilde{\theta} \in (0, \lambda^{-1}]$ is the lease θ satisfying (2.11). We have by the mean value theorem and (2.1) that

$$\begin{aligned} f(x_k + \tilde{\theta}\alpha_k d_k) - f_k &= \int_0^1 g(x_k + t\tilde{\theta}\alpha_k d_k)^T(\tilde{\theta}\alpha_k d_k)dt \\ &= \tilde{\theta}\alpha_k g_k^T d_k + \tilde{\theta}\alpha_k \int_0^1 [g(x_k + t\tilde{\theta}\alpha_k d_k) - g_k]^T d_k dt \\ &\leq \tilde{\theta}\alpha_k g_k^T d_k + \frac{1}{2}L\tilde{\theta}^2\alpha_k^2\|d_k\|^2. \end{aligned} \tag{2.12}$$

Since $\alpha_k g_k^T d_k < 0$ and $\tilde{\theta} < \lambda^{-1}$, we can get from (2.11) and (2.12) that

$$|\alpha_k| \geq c\frac{|g_k^T d_k|}{\|d_k\|^2}, \tag{2.13}$$

where $c = 2\lambda/(L + 2\delta)$. (2.13) also holds if $\alpha_k = 0$. Thus by (2.4) and the definitions of $N_i(k) : i = 1, 2$, (2.5) and (2.6) hold. This completes our proof.

In this paper, we also concern ourselves with the generalized Armijo line search.

Line Search (B). *Given $\lambda \in (0, 1)$ and $\delta \in (0, 1)$. Determine the smallest integer $m \geq 0$ such that, if one defines*

$$\alpha_k = sign\,(-g_k^T d_k)\,\lambda^m, \tag{2.14}$$

then

$$f(x_k + \alpha_k d_k) - f_k \leq \delta\alpha_k g_k^T d_k. \tag{2.15}$$

Similarly to [3], we have the following lemma for the generalized Armijo line search.

Lemma 2.3. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider any method in the form (1.2), where d_k is an arbitrary direction and α_k is obtained by line search (A). Then for every k, we have that either*

$$f_{k+1} - f_k \leq -\eta_1|g_k^T d_k| \tag{2.16}$$

or

$$f_{k+1} - f_k \leq -\eta_2\frac{(g_k^T d_k)^2}{\|d_k\|^2}, \tag{2.17}$$

where η_1 and η_2 are some positive constants.

Proof. For every k, if $|\alpha_k| \in (0, 1)$, we have that

$$f(x_k + \alpha_k d_k) - f_k \leq \delta\alpha_k g_k^T d_k \tag{2.18}$$

and

$$f(x_k + \lambda^{-1}\alpha_k d_k) - f_k \geq \delta\lambda^{-1}\alpha_k g_k^T d_k. \tag{2.19}$$

In this case, similarly to the proof of (2.13) in Lemma 2.2, we can use (2.19), (2.1) and the mean value theorem to prove that

$$|\alpha_k| \geq c\frac{|g_k^T d_k|}{\|d_k\|^2}, \tag{2.20}$$

where $c = \lambda(1-\delta)/L$. Therefore we have either (2.16) holds with $\eta_1 = \delta$ or (2.17) with $\eta_2 = \delta c$.

3. Main Results

Suppose that $g_k \neq 0$ for all k for otherwise a stationary point has been found. Also suppose that $d_k \neq 0$ for all k because it follows from $d_k = 0$ and (1.3) that $d_{k+1} = -g_{k+1}$. Then the method either takes infinite steepest descent steps and hence gives (1.6) or can be regarded to start at $x_{\bar{k}+1}$, where $\bar{k}$ is the largest k such that $d_k = 0$.

Lemma 3.1. *For the FR method (1.2), (1.3) and (1.4), if denoting*

$$t_k = \frac{\|d_k\|^2}{\|g_k\|^4} \quad \text{and} \quad r_k = \frac{-g_k^T d_k}{\|g_k\|^2}, \tag{3.1}$$

we have for all $k \geq 1$,

$$t_k \leq \sum_{i=1}^{k} \frac{r_i^2}{\|g_i\|^2} \tag{3.2}$$

and

$$\sum_{i=1}^{k} \frac{r_i^2}{\|g_i\|^2} \geq \frac{1}{4} \sum_{i=1}^{k} \frac{1}{\|g_i\|^2}. \tag{3.3}$$

Proof. At first, we prove that

$$t_k = -\sum_{i=1}^{k} \frac{1}{\|g_i\|^2} + \sum_{i=1}^{k} \frac{2r_i}{\|g_i\|^2} \tag{3.4}$$

for all $k \geq 1$. Since $d_1 = -g_1$, (3.2) holds for $k = 1$. For $i \geq 2$, it follows from (1.3) that

$$d_i + g_i = \beta_i d_{i-1}. \tag{3.5}$$

Squaring both sides of the above equation, we get that

$$\|d_i\|^2 = -\|g_i\|^2 - 2g_i^T d_i + \beta_i^2 \|d_{i-1}\|^2. \tag{3.6}$$

Dividing (3.6) by $\|g_k\|^4$ and using (1.4) and (3.1),

$$t_i = t_{i-1} - \frac{1}{\|g_i\|^2} + \frac{2r_i}{\|g_i\|^2}. \tag{3.7}$$

Summing this for $i = 2, \ldots, k$ and noting $t_1 = 1/\|g_1\|^2$ and $r_1 = 1$, we see that (3.4) holds for $k \geq 2$. Thus (3.4) holds for all $k \geq 1$.

Since $(1 - r_i)^2 \geq 0$ and hence $r_i^2 \geq 2r_i - 1$, (3.2) follows from (3.4). Noting that $t_k \geq 0$, (3.4) also implies that

$$\sum_{i=1}^{k} \frac{1}{\|g_i\|^2} \leq 2 \sum_{i=1}^{k} \frac{r_i}{\|g_i\|^2}. \tag{3.8}$$

Since $(1 - 2r_i)^2 \geq 0$ and hence $4r_i^2 \geq 4r_i - 1$, it follows that

$$4 \sum_{i=1}^{k} \frac{r_i^2}{\|g_i\|^2} \geq 4 \sum_{i=1}^{k} \frac{r_i}{\|g_i\|^2} - \sum_{i=1}^{k} \frac{1}{\|g_i\|^2} \geq \sum_{i=1}^{k} \frac{1}{\|g_i\|^2}. \tag{3.9}$$

Therefore (3.3) holds, which concludes the proof.

The following lemma is derived from [26].

Lemma 3.2. *Suppose that $\{a_i\}$ and $\{b_i\}$ are positive number sequences. Then if*

$$\sum_{k\geq 1} a_k = \infty \tag{3.10}$$

and for all $k \geq 1$,

$$b_k \leq c_1 + c_2 \sum_{i=1}^{k} a_i, \tag{3.11}$$

where c_1 and c_2 are positive constants, we have that

$$\sum_{k\geq 1} a_k/b_k = \infty. \tag{3.12}$$

Proof. (3.10) implies that for any fixed k,

$$\lim_{l\to\infty} \sum_{i=k+1}^{k+l} a_i = \infty. \tag{3.13}$$

It follows from (3.11) that

$$\sum_{i=k+1}^{k+l} a_i/b_i \geq \sum_{i=k+1}^{k+l} a_i/(c_1 + c_2 \sum_{j=1}^{k+l} a_j). \tag{3.14}$$

Letting $l \to \infty$ in (3.14) and noticing (3.13), we have that

$$\lim_{l\to\infty} \sum_{i=k+1}^{k+l} a_i/b_i \geq 1/c_2. \tag{3.15}$$

Since k in (3.15) is arbitrary, (3.12) holds.

Now we are able to prove the convergence of the FR method using line search (A) under Assumption 2.1 on the objective function.

Theorem 3.3. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the FR method (1.2), (1.3) and (1.4), where the stepsize α_k is obtained by line search (A). Then we have that*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.16}$$

Proof. Since Lemma 2.2 holds, we have by (2.4) and the Cauchy inequality that

$$\sum_{i=1}^{k} \|s_k\| \leq \sqrt{k \sum_{i=1}^{k} \|s_k\|^2} \leq \sqrt{Mk}. \tag{3.17}$$

The Lipschitz condition (2.1) and (1.2) give that

$$\|g_{k+1}\| \leq \|g_k\| + L\|s_k\|. \tag{3.18}$$

Combining (3.17) and (3.18), we have that

$$\sum_{k\geq 1} \frac{1}{\|g_k\|^2} = \infty, \tag{3.19}$$

which with (3.3) indicates that

$$\sum_{k\geq 1} \frac{r_k^2}{\|g_k\|^2} = \infty. \tag{3.20}$$

Suppose that (3.16) is not true. Then there exists a constant $\tau > 0$ such that

$$\|g_k\| \geq \tau, \quad \text{for all } k \geq 1. \tag{3.21}$$

Then it follows from this and (3.2) that

$$\frac{t_k}{\|g_k\|^2} \leq \frac{1}{\tau^2} \sum_{i=1}^{k} \frac{r_i^2}{\|g_i\|^2}. \tag{3.22}$$

Besides it, we have by using $|g_k^T d_k| \leq \|g_k\| \, \|d_k\|$, (2.5) and (3.21) that

$$\lim_{k\to\infty} \sum_{i\in N_1(k)} \frac{r_i^2}{\|g_i\|^2} < \infty, \tag{3.23}$$

which with (3.20) implies that

$$\lim_{k\to\infty} \sum_{i\in N_2(k)} \frac{r_i^2}{\|g_i\|^2} = \infty \tag{3.24}$$

and for all large k,

$$\sum_{i=1}^{k} \frac{r_i^2}{\|g_i\|^2} \leq 2 \sum_{i\in N_2(k)} \frac{r_i^2}{\|g_i\|^2}. \tag{3.25}$$

Thus by (3.22),(3.24) and (3.25), we can apply Lemma 3.2 to get that

$$\lim_{k\to\infty} \sum_{i\in N_2(k)} \frac{r_i^2}{t_i} = \infty. \tag{3.26}$$

The above relation contradicts (2.6). Therefore we have that $\liminf \|g_k\| = 0$.

The convergence of the PRP method using line search (A) for generally problems can be deduced as one corollary of Theorem 3.3 to some extent, as described as follows.

Theorem 3.4. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the PRP method (1.2), (1.3) and (1.5), where the stepsize α_k is obtained by line search (A). Then we have that* $\liminf \|g_k\| = 0$.

Proof. We proceed by contradiction and assume that (3.21) holds. (2.4) implies that

$$\lim_{k\to\infty} \|s_k\| = 0 \tag{3.27}$$

which with (2.1) and (3.21) shows that for all large k,

$$|\beta_k^{PRP}| \leq \frac{\|g_{k+1}\| \, L \, \|s_k\|}{\|g_k\|^2} \leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} = \beta_k^{FR}. \tag{3.28}$$

Since (3.2) and (3.3) still hold for any method (1.2)-(1.3) with β_k such that

$$|\beta_k| \leq \beta_k^{FR} \tag{3.29}$$

for all large k, we can prove (3.16) similarly to Theorem 3.3. (3.16) contradicts (3.21), which indicates that $\liminf \|g_k\| = 0$.

Note that the family of the methods (1.2)-(1.3) with β_k satisfying (3.29) was first introduced in [12]. See [16, 27, 28] for some other families of the methods related to the FR method.

As referred to in Section 1, the FR method using the standard Armijo line search converges globally for generally problems if each search direction has the descent property. Since one can show that (2.3) always holds in the $n = 3$, $m = 6$ example of Powell [24], it seems not possible to establish a similar convergence result for the PRP method. Nevertheless, if the objective function f is uniformly convex, it can be shown that in this case the PRP method converges to the unique minimizer of f.

Theorem 3.5. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the PRP method (1.2), (1.3) and (1.5), where the stepsize α_k is obtained by line search (B). Then if there exists a constant $\eta \geq 0$ such that*

$$[g(x) - g(y)]^T(x - y) \geq \eta \|x - y\|^2, \quad \textit{for all } x, y \in \mathcal{L},, \tag{3.30}$$

we have that $\lim \|g_k\| = 0$.

Proof. By (3.30) and the mean value theorem, we can prove similarly to (2.12) that

$$f(x_k + \alpha_k d_k) - f_k \geq \alpha_k g_k^T d_k + \frac{1}{2}\eta\alpha_k^2\|d_k\|^2. \tag{3.31}$$

This and (2.15) imply that

$$\|s_k\|^2 \leq -c\alpha_k g_k^T d_k, \tag{3.32}$$

where $c = 2(1 - \delta)/\eta$. Since the limit of f_k exists, we have from (2.15) that

$$\lim_{k\to\infty} \alpha_k g_k^T d_k = 0. \tag{3.33}$$

Thus (3.27) follows from (3.32) and (3.33). Therefore similarly to the proof of Theorem 3.4, we can prove that $\lim \|g_k\| = 0$, which concludes the proof.

4. Discussion

We have shown that a simplified Armijo-type line search can ensure the global convergences of the FR method and the PRP method for generally problems. To deal with the case when a descent direction is not produced, the generalized line search technique is used in this paper. One can prove that the next directions produced by both the FR method and the PRP method have the descent property if the stepsize α_k is sufficient small. Similar convergence result can be established all the same.

The convergence properties of the FR method and the PRP method under the simplified Armijo-type line search are satisfactory. But it is not yet known if similar convergence results can be established for other nonlinear conjugate gradient methods.

References

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

[2] L. Armijo, Minimization of functions having Lipschitz continuous first partial derivatives, *Pacific Journal of Mathematics,* **16**:1 (1966), 1-3.

[3] R.H. Byrd, J. Nocedal, A Tool for the Analysis of Quasi-Newton Methods with Application To Unconstrained Minimization, *SIAM J. Numer. Anal.,* **26**:3 (1989), 727-739.

[4] Y.H. Dai, Analyses of Nonlinear Conjugate Gradient Methods, Ph.D. thesis, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1997. (in Chinese)

[5] Y.H. Dai, Further Insight into the Convergence of the Fletcher-Reeves Method, ICM-97-50, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1997.

[6] Y.H. Dai, Y. Yuan, A Nonlinear Conjugate Gradient Method with Nice Global Convergence Properties, Research report ICM-95-038, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1995.

[7] Y.H. Dai, Y. Yuan, Convergence properties of the Fletcher-Reeves method, *IMA J. Numer. Anal.,* **16**:2 (1996).

[8] Y.H. Dai, Y. Yuan, Convergence of the Fletcher-Reeves Method under A Generalized Wolfe Search, *Journal Computational Mathematics of Chinese Universities,* **2** (1996).

[9] Daniel, The conjugate gradient method for linear and nonlinear operators, *SIAM J. Numer. Anal.,* **4** (1967), 10-26.

[10] R. Fletcher, Practical Methods of Optimization vol. 1: Unconstrained Optimization, John Wiley & Sons (New York), 1987.

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

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

[13] L. Grippo, F. Lampariello, S. Lucidi, Global convergence and stabilization of unconstrained minimization methods without derivatives, *Journal of Optimization Theory and Applications,* **56** (1988) 385-406.

[14] L. Grippo, S. Lucidi, A globally convergent version of the Polak-Ribière conjugate gradient method, Dipartimento di Informatica e Sistermatica, Universita degli Studi fi Roma "Ra Sapeinza", R. 08-95, 1995. (to appear in: Math. Prog.)

[15] M.R. Hestenes, E.L. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Standards Sect.,* **5**:49 (1952), 409-436.

[16] Y.F. Hu, C. Storey, Global convergence of the Fletcher-Reeves methods, *JOTA,* **71**:2 (1991), 399-405.

[17] R. De Leone, M. Gaudioso, L. Grippo, Stopping criteria for linesearch methods without derivatives, Mathematical Programming 30, (1984) 285-300.

[18] Y. Liu, C. Storey, Efficient generalized conjugate algorithm, Part 1: Theory, *Journal of Optimization Theory and Applications,* **69**:1 (1991), 129-137.

[19] G.H. Liu, J.Y. Han, H.X. Yin, Global convergence of the Fletcher-Reeves algorithm with an inexact line search, Report, Institute of Applied Mathematics, Chinese Academy of Sciences, 1993.

[20] J.L. Nazareth, A view of conjugate gradient-related algorithms for nonlinear optimization, in: Linear and Nonlinear Conjugate Gradient-Related Methods, L. Adams and J. L. Nazareth, eds., SIAM, Philadelphia, 1996, 149-163.

[21] J. Nocedal, Theory of algorithm for unconstrained optimization, *Acta Numerica* (1991), 199-242.

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

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

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

[25] M.J.D. Powell, Convergence Properties of Algorithms for Nonlinear Optimization, *SIAM Review,* **28** (1986), 487-500.

[26] D. Pu, W. Yu, On the convergence properties of the DFP algorithms, Annals of Operations Research 24 (1990) 175-184.

[27] D. Touati-Ahmed, C. Storey, Efficient hybrid conjugate gradient techniques, *J. Optimization Theory Appl.,* **64** (1990), 379-397.

[28] C.Y. Wang, Y.Z. Zhang, Global convergence properties of s-related conjugate gradient methods, Report, Qufu Normal University, 1996.

[29] P. Wolfe, Convergence conditions for ascent methods, *SIAM Review,* **11** (1969), 226-235.

[30] G. Zoutendijk, Nonlinear Programming, Computational Methods, in: Integer and Nonlinear Programming (J. Abadie, ed.), 1970, pp. 37-86.