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.