Journal of Computational Mathematics, Vol.20, No.6, 2002, 575–582.

A NOTE ON THE NONLINEAR CONJUGATE GRADIENT METHOD *1)

Y.H. Dai Y. Yuan
(*State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, P. O. Box 2719, Beijing 100080, China*)

Abstract

The conjugate gradient method for unconstrained optimization problems varies with a scalar. In this note, a general condition concerning the scalar is given, which ensures the global convergence of the method in the case of strong Wolfe line searches. It is also discussed how to use the result to obtain the convergence of the famous Fletcher-Reeves, and Polak-Ribiére-Polyak conjugate gradient methods. That the condition cannot be relaxed in some sense is mentioned.

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

1. Introduction

The conjugate gradient method is highly useful for minimizing a smooth function of n variables,

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

especially when n is large. It has the following form

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

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

where $g_k = \nabla f(x_k)$, α_k is a stepsize obtained by a one-dimensional line search and β_k is a scalar. Because α_k is not the exact one-dimensional minimizer in practice and f is not a quadratic, many formulas have been proposed to compute the scalar β_k. Two well-known formulas for β_k are called the Fletcher-Reeves (FR), and Polak-Ribiére-Polyak (PRP) formulas (see [8, 16, 17]). They are given by

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

and

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

respectively, where $\|\cdot\|$ means the Euclidean norm. See Dai & Yuan [2], Daniel [6], Fletcher [7], Hestenes & Stiefel [11], Liu & Storey [13] *etc.* for other formulas of β_k.

In recent years, many authors studied the nonlinear conjugate gradient method especially from the angle of global convergence. Because its properties can be very different with the choice of β_k (see Powell [14]), the nonlinear conjugate gradient method was often analyzed

* Received.
1) Research partially supported by Chinese NSF grants 19525101, 19731010, 19801033, and 10171104

individually, for example, see Al-Baali [1], Dai & Yuan [3, 4], Gilbert & Nocedal [9], Grippo & Lucidi [10], Liu *et al* [12], Powell [15], Qi [19] and Touati-Ahmed & Storey [20]. Dai *et al* [5] studied the general conjugate gradient method in the absence of the sufficient descent condition and proposed a sufficient condition ensuring the global convergence (see also Lemma 2.3). Since the nonlinear conjugate gradient method varies with the choice of β_k, we wonder what condition on β_k guarantees the convergence of the method.

This paper is organized as follows. After giving some preliminaries in the next section, we will prove in Section 3 that a mild condition on β_k results in global convergence of the nonlinear conjugate gradient method in the case of strong Wolfe line searches. Section 4 discusses how to use the result to obtain the convergence of the famous FR, and PRP conjugate gradient method. In the last section, it is mentioned that the condition on β_k cannot be relaxed in some sense.

2. Preliminaries

For convenience, we assume that $g_k \neq 0$ for all k, for otherwise a stationary point has been found. We give the following basic assumption on the objective function.

Assumption 2.1. *(i) The level set* $\mathcal{L} = \{x \in \Re^n : f(x) \leq f(x_1)\}$ *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*

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

Assumption 2.1 implies that there exists a constant $\overline{\gamma}$ such that

$$\|g(x)\| \leq \overline{\gamma}, \qquad \text{for all } x \in \mathcal{L}. \tag{2.2}$$

The stepsize α_k in (1.2) is computed by carrying out certain line searches. The Wolfe line search [21] is to find a positive stepsize α_k such that

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{2.3}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{2.4}$$

where $0 < \delta < \sigma < 1$. Under Assumption 2.1 on f, we state the following result, which was essentially obtained by Zoutendijk [23] and Wolfe [21, 22].

Lemma 2.2. *Suppose that* x_1 *is a starting point for which Assumption 2.1 holds. Consider any iterative method (1.2), where* d_k *is a descent direction and* α_k *is computed by the standard Wolfe line search. Then*

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

In the convergence analysis and implementation of conjugate gradient methods, the stepsize α_k is often computed by the strong Wolfe line search, namely, (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 $0 < \delta < \sigma < 1$. Dai *et al* [5] proved the following general convergence result for any conjugate gradient method using the strong Wolfe line search.