Computers & Operations Research 40 (2013) 2656–2661

Contents lists available at SciVerse ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

# Convergence properties of a class of nonlinear conjugate gradient methods

Jinkui Liu *,1

*School of Mathematics and Statistics, Chongqing Three Gorges University, Chongqing, China*

ARTICLE INFO

Available online 1 June 2013

*Keywords:*
Unconstrained optimization
Conjugate gradient method
Strong Wolfe line search
Descent property
Convergence property

ABSTRACT

Conjugate gradient methods are a class of important methods for unconstrained optimization problems, especially when the dimension is large. In this paper, we study a class of modified conjugate gradient methods based on the famous LS conjugate gradient method, which produces a sufficient descent direction at each iteration and converges globally provided that the line search satisfies the strong Wolfe condition. At the same time, a new specific nonlinear conjugate gradient method is constructed. Our numerical results show that the new method is very efficient for the given test problems by comparing with the famous LS method, PRP method and CG-DESCENT method.



## 1. Introduction

In the reference [2], Liu and Storey propose the famous LS nonlinear conjugate gradient method (called LS method). The important property of the LS method is that it has good numerical results. But its global convergence has not been thoroughly proved under the Wolfe-type line search condition. The purpose of this paper is to study a class of conjugate gradient methods related to the famous LS method.

Consider the following unconstrained optimization problem:

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

where $f(x)$ is smooth and its gradient $g(x)$ is available. Conjugate gradients methods are efficient to solve the above problems, and have the following iteration form:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.1}$$

$$d_k = \begin{cases} -g_k, & \text{for } k=1; \\ -g_k + \beta_k d_{k-1}, & \text{for } k\geq 2, \end{cases} \tag{1.2}$$

where $g_k = -\nabla f(x_k)$, $\alpha_k > 0$ is a step length determined by some line search; $d_k$ is the search direction and $\beta_k$ is a scalar. The formula of $\beta_k$ should be so chosen that the method reduces to the linear conjugate gradient method in some case when $f(x)$ is strictly convex and the line search is exact.

Some well-known formulas of $\beta_k$ are called the Fletcher–Reeves (FR) [1], Liu–Story (LS) [2], Polak–Ribière–Polyak (PRP) [3,4] and Hager–Zhang (HZ) [5] formulas, i.e.,

$$\beta_k^{FR} = \frac{||g_k||^2}{||g_{k-1}||^2}, \ [1]; \beta_k^{LS} = -\frac{g_k^T y_{k-1}}{d_{k-1}^T g_{k-1}}, \ [2]; \beta_k^{PRP} = \frac{g_k^T y_{k-1}}{||g_{k-1}||^2}, \ [3,4];$$

$$\beta_k^{HZ} = \left(y_{k-1} - 2d_{k-1}\frac{||y_{k-1}||^2}{d_{k-1}^T y_{k-1}}\right)^T \frac{g_k}{d_{k-1}^T y_{k-1}}, \ [5];$$

where $\|\cdot\|$ is the Euclidean norm and $y_{k-1} = g_k - g_{k-1}$. Their corresponding conjugate gradient methods were generally specified as FR, LS, PRP and HZ methods. Obviously, if $f$ is a strictly convex quadratic function and the line search is exact, the above methods are equivalent.

In the past few years, the LS and PRP methods have been regarded as the most efficient conjugate gradient method in practical computation, which makes them research widely, see [2–4,6–11]. Hager and Zhang [5] discussed the global convergence of the HZ method for strong convex functions under the Wolfe line search, i.e., $\alpha_k$ satisfies

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \tag{1.3}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{1.4}$$

where $0 < \delta < \sigma < 1$. In order to prove the global convergence for general functions, Hager and Zhang modified the parameter $\beta_k^{HZ}$ as

$$\beta_k^{MHZ} = \max\{\beta_k^{HZ}, \eta_k\} \tag{1.5}$$

where $\eta_k = (-1/||d_{k-1}||\min\{\eta, ||g_{k-1}||\})$, $\eta > 0$. The corresponding method of (1.5) is the famous CG-DESCENT method.

Gilbert and Nocedal [12] investigate global convergence properties of the dependent FR method with $\beta_k$ satisfying $|\beta_k| \leq |\beta_k^{FR}|$, provided that the line search satisfies the strong Wolfe conditions,

* Tel.: +86 13658363032.
*E-mail address:* liujinkui2006@126.com

[1] This work was supported by The Nature Science Foundation of Chongqing Education Committee (KJ121112).

0305-0548/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cor.2013.05.013

i.e., $\alpha_k$ satisfies (1.3) and

$$|g(x_k+\alpha_k d_k)^T d_k| \le -\sigma g_k^T d_k, \tag{1.6}$$

where $0<\delta<\sigma<1$.

The above observation motivates us to construct a class of conjugate gradient methods in which $\beta_k$ satisfies

$$\left|\beta_k\right| \le t_k\left|\beta_k^{LS}\right|, \tag{1.7}$$

where

$$t_k=\frac{u}{\sigma(1+\mu_k)},\quad \mu_k=\frac{|g_k^T g_{k-1}|}{||g_k||^2},\quad 0<u<\frac{1}{2} \text{ and } \sigma\in(0,1)$$

In Section 2, we will investigate global convergence properties of the new modified conjugate gradient method with the strong Wolfe line search. In Section 3, we will give a specific nonlinear conjugate gradient method which originates in the new modified conjugate gradient method, and some numerical results are also reported.

## 2. The main results

In this section, we always assume that $||g_k||\neq 0$ for all $k$, for otherwise a stationary point has already been found. At the same time, in order to guarantee the global convergence of the new method, we make the following assumption on the objective function $f(x)$.

**Assumption (H).**

(i) The level set $\Omega=\{x\in R^n | f(x)\le f(x_1)\}$ is bounded, where $x_1$ is the starting point.
(ii) In some neighborhood V of $\Omega$, $f$ is differentiable and its gradient g is Lipchitz continuous, namely, there exists a constant $L>0$ such that

$$||g(x)-g(y)|| \le L||x-y||,\quad \text{for all } x,y\in V.$$

Obviously, from Assumption (H), we know that there exists a constant $\tilde{r}>0$, such that

$$||g(x)||\le \tilde{r}\quad \text{for all } x\in V. \tag{2.1}$$

**Lemma 2.1.** *Consider any method* (1.1)–(1.2), *where $\beta_k$ satisfies* (1.7) *and $\alpha_k$ satisfies the strong Wolfe line search* (1.3) *and* (1.6). *Then*

$$g_k^T d_k \le -(1-u)||g_k||^2. \tag{2.2}$$

**Proof.** We prove the conclusion by induction. Since $||g_1||^2=-g_1^T d_1$ and $u\in(0,1/2)$, the conclusion (2.2) holds for $k=1$. Now we assume that the conclusion is true for $k-1$ and $g_k\neq 0$, then $g_{k-1}^T d_{k-1}<0$. We need to prove that the result holds for $k$.

Multiplying (1.2) by $g_k^T$, we have

$$g_k^T d_k = -||g_k||^2+\beta_k g_k^T d_{k-1}.$$

Then from (1.7) and (1.6), we get

$$\begin{aligned} g_k^T d_k &\le -\left|\left|g_k\right|\right|^2+|\beta_k|\cdot|g_k^T d_{k-1}| \le -||g_k||^2+t_k\cdot|\beta_k^{LS}|\cdot|g_k^T d_{k-1}| \\ &\le -\left|\left|g_k\right|\right|^2+\frac{u||g_k||^2}{\sigma(||g_k||^2+|g_k^T g_{k-1}|)}\cdot\frac{||g_k||^2+|g_k^T g_{k-1}|}{-g_{k-1}^T d_{k-1}}\cdot(-\sigma g_{k-1}^T d_{k-1}) \\ &\le -(1-u)||g_k||^2. \end{aligned}$$

From above inequality, the conclusion (2.2) also holds.

According to (2.2) and $0<u<\frac{1}{2}$, we also have

$$||g_k||^2 < -2g_k^T d_k. \tag{2.3}$$

We state a general convergence result as follows. This result was essentially proved by Zoutendijk [13]. It is important in the convergence analyses of nonlinear optimization methods.

**Lemma 2.2.** *Suppose Assumption (H) holds. Consider any method* (1.1)–(1.2), *where $d_k$ satisfies $g_k^T d_k<0$ for $k\in N^+$ and $\alpha_k$ satisfies the Wolfe line search* (1.3) *and* (1.4). *Then*

$$\sum_{k\ge 1}\frac{(g_k^T d_k)^2}{||d_k||^2} < +\infty. \tag{2.4}$$

The strong Wolfe line search is a special case of the Wolfe line search, so the Lemma 2.2 also holds under the strong Wolfe line search.

**Lemma 2.3.** *Consider any method* (1.1)–(1.2), *where $\beta_k$ satisfies* (1.7) *and $\alpha_k$ satisfies the strong Wolfe line search* (1.3) *and* (1.6). *If there exists a constant $r>0$, for $\forall k\ge 1$ such that*

$$||g_k||\ge r. \tag{2.5}$$

Then we have

$$\sum_{k\ge 2}||u_k-u_{k-1}||^2 < +\infty.$$

where

$$u_k=\frac{d_k}{||d_k||}.$$

**Proof.** From (2.2), we know that $d_k\neq 0$, $\forall k\in N^+$. Define the quantities

$$r_k=\frac{-g_k}{||d_k||} \text{ and } \delta_k=\frac{\beta_k||d_{k-1}||}{||d_k||}$$

In the following, we first prove that $1+\delta_k\neq 0$ holds. Obviously, the inequality holds when the parameter $\beta_k\ge 0$. In order to prove that the inequality also holds when $\beta_k<0$, we apply the contradiction. Suppose that $1+\delta_k=0$ holds, i.e., $-\beta_k||d_{k-1}||=||d_k||$.

From $\beta_k<0$, then we have

$$||\beta_k d_{k-1}||=||d_k||$$

From (1.2), we have $||d_k+g_k||=||\beta_k d_{k-1}||$, therefore,

$$||d_k+g_k||=||d_k||$$

Squaring the equality, we get

$$||g_k||^2=-2g_k^T d_k,$$

which is contradictive with (2.3). Hence, $1+\delta_k\neq 0$ always holds when $\beta_k$ satisfies (1.7) under the strong Wolfe line search, i.e., there exists a constant $\rho>0$ such that

$$|1+\delta_k|\ge\rho. \tag{2.6}$$

From (1.2), we have

$$u_k=\frac{d_k}{||d_k||}=\frac{-g_k+\beta_k d_{k-1}}{||d_k||}=r_k+\delta_k u_{k-1}.$$

According to the definition of $u_k$, we have $||u_k||=1$, then

$$||r_k||=||u_k-\delta_k u_{k-1}||=||\delta_k u_k-u_{k-1}||. \tag{2.7}$$

By (2.6) and (2.7), we have

$$\begin{aligned} ||u_k-u_{k-1}|| &= \frac{1}{|1+\delta_k|}||(1+\delta_k)\cdot(u_k-u_{k-1})|| \\ &\le \frac{1}{|1+\delta_k|}(||u_k-\delta_k u_{k-1}||+||\delta_k u_k-u_{k-1}||) \\ &\le \frac{2}{\rho}||r_k||. \end{aligned} \tag{2.8}$$

From (2.2), (2.4), (2.5) and (2.8), we have

$$\frac{\rho^2 r^2(1-u)^2}{4}\sum_{k\ge 1}||u_k-u_{k-1}||^2 \le (1-u)^2\sum_{k\ge 1}(||g_k||^2\cdot r_k^2)$$