Journal of Computational and Applied Mathematics 305 (2016) 92–114

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Some modified conjugate gradient methods for unconstrained optimization☆

Xuewu Du [a], Peng Zhang [b,a,*], Wenya Ma [a]

[a] School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, PR China

[b] School of Management, Shanghai University, Shanghai 200444, PR China

ARTICLE INFO

Article history:
Received 8 August 2015
Received in revised form 26 March 2016

MSC:
90C30
90C25

Keywords:
Unconstrained optimization
Conjugate gradient method
Line search
Sufficient descent property
Global convergence

ABSTRACT

Conjugate gradient methods are highly useful for solving large scale optimization problems because they do not require the storage of any matrices. Motivated by the construction of conjugate gradient parameters in some existing conjugate gradient methods, we propose four modified conjugate gradient methods, named NVLS, NVPRP*, NVHS* and NVLS* respectively, and prove that these methods with the strong Wolfe line search possess sufficient descent property, and are globally convergent when the parameter in line search conditions is restricted in some suitable interval. Preliminary numerical results show that the NVPRP*, NVHS* and NVLS* methods are more efficient than many existing conjugate gradient methods for a large number of test problems from a CUTEr collection.

1. Introduction

Consider the unconstrained optimization problem

$$\min f(x), \quad x \in R^n, \tag{1}$$

where $f : R^n \to R$ is a continuously differentiable function, and its gradient g is available.

Conjugate gradient methods are a class of important iterative methods for solving (1), especially if the dimension n is large. They are of the form

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

where $\alpha_k > 0$ is a stepsize obtained by some line search technique, d_k is the search direction defined by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0; \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{3}$$

where g_k denotes the gradient $g(x_k)$ of f at x_k, the scalar β_k, called *conjugate gradient parameter*, is so chosen that (2)–(3) reduces to the *linear* conjugate gradient method in the case when $f(x)$ is a strictly convex quadratic function, and α_k is the exact one-dimensioned minimizer.

☆ This work is partially supported by the Natural Science Foundation of China (Grant 11171363) and the Natural Science Foundation of Chongqing (Grants cstc2015jcyjBX0029, cstc2013jjB00001, cstc2013jcyjA00021).

* Corresponding author at: School of Management, Shanghai University, Shanghai 200444, PR China. Tel.: +86 023 65362798; fax: +86 023 65362798.
E-mail address: zhangpeng_9001@126.com (P. Zhang).

http://dx.doi.org/10.1016/j.cam.2016.04.004
0377-0427/

Some of the well-known nonlinear conjugate gradient methods are the Fletcher–Reeves (FR) [1], Polak–Ribière–Polyak (PRP) [2,3], Hestenes–Stiefel (HS) [4], Conjugate Descent (CD) [5], Liu–Storey (LS) [6] and Dai–Yuan (DY) [7] methods in which formulas for β_k are given, respectively, by

$$\beta_k^{\mathrm{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \beta_k^{\mathrm{PRP}} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \qquad \beta_k^{\mathrm{HS}} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}},$$

$$\beta_k^{\mathrm{CD}} = \frac{\|g_k\|^2}{-g_{k-1}^T d_{k-1}}, \qquad \beta_k^{\mathrm{LS}} = \frac{g_k^T y_{k-1}}{-g_{k-1}^T d_{k-1}}, \qquad \beta_k^{\mathrm{DY}} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}},$$

where $\|\cdot\|$ denotes the Euclidean norm, and $y_{k-1} = g_k - g_{k-1}$.

Some convergence analyses of conjugate gradient methods require the stepsize α_k to be determined by the exact line search. However, since exact line search is usually expensive and impractical, inexact line searches are often used in convergence analyses and implementations of conjugate gradient methods. Inexact line searches include the Armijo, Goldstein, strong Wolfe, Wolfe, generalized Wolfe line searches and their variants. The strong Wolfe line search (SWLS) is to find an α_k satisfying

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k \tag{4}$$

and

$$|g(x_k + \alpha_k d_k)^T d_k| \le -\sigma g_k^T d_k, \tag{5}$$

where $0 < \delta < \sigma < 1$. The Wolfe line search (WLS) is to find an α_k satisfying (4) and

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{6}$$

where again $0 < \delta < \sigma < 1$. The generalized Wolfe line search (GWLS), which was originally introduced by Dai and Yuan [8], is to find an α_k satisfying (4) and

$$\sigma_1 g_k^T d_k \le g(x_k + \alpha_k d_k)^T d_k \le -\sigma_2 g_k^T d_k, \tag{7}$$

where $0 < \delta < \sigma_1 < 1$ and $\sigma_2 \ge 0$. Obviously, (5) can be viewed as a special case of (7) with $\sigma_1 = \sigma_2 = \sigma$, and (6) as a special case of (7) with $\sigma_1 = \sigma$ and $\sigma_2 = +\infty$.

A requirement for an optimization method to use the above line searches is that, all search directions must have the descent property, that is,

$$g_k^T d_k < 0 \tag{8}$$

for all k. In this paper, we say that, a conjugate gradient method is *descent* or possesses the *descent property* if the descent condition (8) holds for all k, and it is *sufficient descent* or possesses the *sufficient descent property* if there exists some constant $c > 0$ such that the sufficient descent condition

$$g_k^T d_k \le -c\|g_k\|^2 \tag{9}$$

holds for all k. The sufficient descent condition (9) is often used to analyze the global convergence of nonlinear conjugate gradient methods.

To simplify the statements of the theoretical analyses and results, we always assume that $g_k \ne 0$ for all k, for otherwise a stationary point has been found. Assume also that $\beta_k \ne 0$ for all k. This is because if $\beta_k = 0$, the direction d_k in (3) reduces to the negative gradient direction. Thus, either the method converges globally if $\beta_k = 0$ for infinite number of k, or one can take some x_k as the new initial point. In addition, we say that a method is *globally convergent* if

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{10}$$

In practical computations, if the FR method generates a bad direction and a tinny step from x_{k-1} to x_k, the next direction and the next step are also likely to be poor unless a restart along the negative gradient direction is performed (see [9]). In spite of such a drawback, it has been shown that the FR method has nice theoretical convergence properties (see [10–13,8]). In addition, by using the technique in [8], Du and Xu [14] investigated the methods of the form (2)–(3) with β_k satisfying

$$0 \le \beta_k \le \beta_k^{\mathrm{FR}}. \tag{11}$$

They showed that any method (2)–(3) with β_k satisfying (11) possesses the descent property, and is globally convergent if the GWLS (4) and (7) with $\sigma_1 + \sigma_2 \le 1$ is used. The numerical performances of the CD and DY methods are very similar to the FR method since the scalar β_k in these two methods have the same numerator as the FR method.

The PRP method performs much better than the FR method for many optimization problems because it can automatically recover once a small step is generated (see [9]). The numerical performances of the HS and LS methods are very similar to the PRP method since the scalar β_k in these methods have the same numerator. Nevertheless, the theoretical convergence