Applied Numerical Mathematics 92 (2015) 70–81

Contents lists available at ScienceDirect

Applied Numerical Mathematics

www.elsevier.com/locate/apnum

A three-term conjugate gradient algorithm for large-scale unconstrained optimization problems ☆

Songhai Deng, Zhong Wan *

School of Mathematics and Statistics, Central South University, Changsha, China

A R T I C L E   I N F O

Article history:
Received 29 December 2013
Received in revised form 10 October 2014
Accepted 28 January 2015
Available online 7 February 2015

Keywords:
Large-scale problems
Three-term conjugate gradient method
Global convergence
Inexact line search
Algorithm

A B S T R A C T

In this paper, a three-term conjugate gradient algorithm is developed for solving large-scale unconstrained optimization problems. The search direction at each iteration of the algorithm is determined by rectifying the steepest descent direction with the difference between the current iterative points and that between the gradients. It is proved that such a direction satisfies the approximate secant condition as well as the conjugacy condition. The strategies of acceleration and restart are incorporated into designing the algorithm to improve its numerical performance. Global convergence of the proposed algorithm is established under two mild assumptions. By implementing the algorithm to solve 75 benchmark test problems available in the literature, the obtained results indicate that the algorithm developed in this paper outperforms the existent similar state-of-the-art algorithms.

1. Introduction

Since there exist many large-scale optimization problems in engineering planning, shape design and structural optimization, it is fundamentally important to develop an efficient algorithm to solve the following basic type of large-scale problems:

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable function such that its gradient is available (see [16,23,27–31] and the references therein). Let $g : \mathbb{R}^n \to \mathbb{R}^n$ denote the gradient function of f, and let g_k denote the value of g at x_k. If the dimension n of problem (1) is large, then the second order information of f is not suitable to be utilized to design a powerful algorithm for solving (1), as in the quasi-Newton method. In this article, we focus on the development of an algorithm to solve (1), which is only involved with the first-order information of f.

With an arbitrarily chosen initial point $x_0 \in \mathbb{R}^n$, an approximate solution sequence of (1), $\{x_k\}$, is often generated by

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

☆ This research is supported by the Natural Science Foundation of Hunan Province (13JJ3002, 14JJ2003) and the National Natural Science Foundation of China (grant No. 71221061, 61375063).

* Corresponding author.
E-mail addresses: dsonghai@163.com (S. Deng), wanmath@163.com (Z. Wan).

http://dx.doi.org/10.1016/j.apnum.2015.01.008
0168-9274/© 2015 IMACS. Published by Elsevier B.V. All rights reserved.

where $k \geq 0$, $d_k \in \mathbb{R}^n$ is called a search direction at x_k and $\alpha_k \geq 0$ is a step size along d_k obtained by some line search rule. In the classical conjugate gradient methods, d_k is given by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if } k > 0. \end{cases} \tag{2}$$

In (2), β_k is called the conjugate parameter. With a different choice of β_k, the obtained method has distinct numerical performance. Denote $y_k = g_{k+1} - g_k$. We present some popular conjugate parameters as follows:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \quad \text{(Hestenes and Stiefel [21], 1952)},$$

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \quad \text{(Fletcher and Reeves [18], 1964)},$$

$$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k} \quad \text{(Dai and Yuan [14], 1999)},$$

$$\beta_k^{HZ} = \frac{1}{d_k^T y_k}\left(y_k - 2d_k \frac{\|y_k\|^2}{d_k^T y_k}\right) g_{k+1} \quad \text{(Hager and Zhang [19], 2005)}.$$

It is proved in [19] that the constructed d_k satisfies $g_k^T d_k \leq -\frac{7}{8}\|g_k\|^2$ if $d_k^T y_k \neq 0$. In virtue of an approximate Wolfe line search, Hager and Zhang developed an efficient and very famous algorithm, called CG_DESCENT, in [20].

To improve the efficiency of the classical conjugate gradient method, a type of three-term conjugate gradient methods has been widely studied recently. The first general three-term conjugate gradient method was proposed by E.M.L. Beale in [9], where by incorporating a restart direction d_t ($t \leq k-1$), the search direction was determined by

$$d_{k+1} = -g_{k+1} + \beta_k d_k + \gamma_k d_t. \tag{3}$$

In Beale algorithm [9], the parameter β_k can be given by β_k^{HS}, β_k^{FR}, or β_k^{DY}, etc., and the parameter

$$\gamma_k = \begin{cases} 0, & k = t+1, \\ \dfrac{g_{k+1}^T y_t}{d_t^T y_t}, & k > t+1. \end{cases} \tag{4}$$

In [25], Nazareth presented another kind of three-term conjugate gradient method, where the search direction is computed by

$$d_{k+1} = -y_k + \frac{y_k^T y_k}{y_k^T d_k} d_k + \frac{y_{k-1}^T y_k}{y_{k-1}^T d_{k-1}} d_{k-1} \tag{5}$$

with $d_{-1} = 0$, $d_0 = 0$. It is proved that if f is a convex quadratic objective function, then for any stepsize α_k, the search directions generated by (5) are conjugate in the coefficient matrix of the quadratic term in f. In [35], a descent modified PRP conjugate gradient algorithm was developed, where the search direction is obtained by the following three-term formula:

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T y_k}{g_k^T g_k} d_k - \frac{g_{k+1}^T d_k}{g_k^T g_k} y_k.$$

In [36], the HS conjugate gradient method was modified by a descent three-term conjugate gradient method. It reads

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T y_k}{s_k^T y_k} s_k - \frac{g_{k+1}^T s_k}{s_k^T y_k} y_k,$$

where $s_k = x_{k+1} - x_k$.

Different from the conventional conjugate gradient methods, the above three-term methods have a remarkable property that the constructed direction is sufficiently descent, namely for each $k \geq 0$, it satisfies that $g_k^T d_k \leq -c\|g_k\|^2$, where c is a given constant.

Very recently, Andrei in [5–7] investigated the following three types of descent three-term gradient methods:

$$d_{k+1} = -\frac{y_k^T s_k}{\|g_k\|^2} g_{k+1} + \frac{y_k^T g_{k+1}}{\|g_k\|^2} s_k - \frac{s_k^T g_{k+1}}{\|g_k\|^2} y_k, \tag{6}$$

$$d_{k+1} = -g_{k+1} - \left(\left(1 + \frac{\|y_k\|^2}{y_k^T s_k}\right) \frac{s_k^T g_{k+1}}{y_k^T s_k} - \frac{y_k^T g_{k+1}}{y_k^T s_k}\right) s_k - \frac{s_k^T g_{k+1}}{y_k^T s_k} y_k, \tag{7}$$