Journal of Computational and Applied Mathematics 261 (2014) 172–182

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Two modified scaled nonlinear conjugate gradient methods

Saman Babaie-Kafaki *

Department of Mathematics, Faculty of Mathematics, Statistics and Computer Sciences, Semnan University, P.O. Box: 35195-363, Semnan, Iran

ARTICLE INFO

Article history:
Received 4 December 2011
Received in revised form 31 May 2013

MSC:
65K05
90C53
49M37

Keywords:
Unconstrained optimization
Scaled nonlinear conjugate gradient method
BFGS update
Modified secant equation
Sufficient descent condition
Global convergence

ABSTRACT

Following the scaled conjugate gradient methods proposed by Andrei, we hybridize the memoryless BFGS preconditioned conjugate gradient method suggested by Shanno and the spectral conjugate gradient method suggested by Birgin and Martínez based on a modified secant equation suggested by Yuan, and propose two modified scaled conjugate gradient methods. The interesting features of our methods are applying the function values in addition to the gradient values and satisfying the sufficient descent condition for the generated search directions which leads to the global convergence for uniformly convex functions. Numerical comparisons between the implementations of one of our methods which generates descent search directions for general functions and an efficient scaled conjugate gradient method proposed by Andrei are made on a set of unconstrained optimization test problems from the CUTEr collection, using the performance profile introduced by Dolan and Moré.

1. Introduction

Conjugate gradient (CG) methods comprise a class of unconstrained optimization algorithms characterized by low memory requirements and strong global convergence properties [1,2]. Although CG methods are not the fastest or most robust optimization algorithms for nonlinear problems available today, they remain very popular for engineers and mathematicians engaged in solving large-scale problems in the following form:

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth nonlinear function and its gradient is available. The iterative formula of a CG method is given by

$$x_0 \in \mathbb{R}^n, \qquad x_{k+1} = x_k + s_k, \qquad s_k = \alpha_k d_k, \quad k = 0, 1, 2, \ldots, \tag{1.2}$$

where α_k is a steplength to be computed by a line search procedure and d_k is the search direction defined by

$$d_0 = -g_0, \qquad d_{k+1} = -g_{k+1} + \beta_k d_k, \quad k = 0, 1, 2, \ldots, \tag{1.3}$$

in which $g_k = \nabla f(x_k)$ and β_k is a scalar called the CG (update) parameter. The steplength α_k is usually chosen to satisfy certain line search conditions [3,4]. Among them, the so-called Wolfe line search conditions [5,6] have attracted special

* Tel.: +98 231 3354160.
E-mail addresses: sbk@semnan.ac.ir, babaiekafaki@gmail.com.

0377-0427/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cam.2013.11.001

attention in convergence analysis and implementation of the CG algorithms, requiring that

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k \nabla f(x_k)^T d_k, \tag{1.4}$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f(x_k)^T d_k, \tag{1.5}$$

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

Different choices for the CG parameter in (1.3) lead to different CG methods (see [7] and the references therein). Although the CG methods are often equivalent in the linear case, that is, when f is a strictly convex quadratic function and α_k is computed by an exact line search, their behavior for general functions may be quite different [8–10].

The interesting feature of linear CG methods leading to the important n-step termination property [3,4] is generation of the search directions $\{d_k\}_{k\geq 0}$ satisfying

$$d_i^T H d_j = 0, \quad \forall i \neq j, \tag{1.6}$$

where H is the Hessian matrix of the objective function. For a general nonlinear function f, from the mean-value theorem there exists some $\xi \in (0, 1)$ such that

$$d_{k+1}^T (g_{k+1} - g_k) = \alpha_k d_{k+1}^T \nabla^2 f(x_k + \xi \alpha_k d_k) d_k.$$

Therefore, in order to compute the CG parameter β_k for a nonlinear CG method in the form of (1.2)–(1.3), it is reasonable to replace (1.6) with the following conjugacy condition:

$$d_{k+1}^T y_k = 0, \tag{1.7}$$

where $y_k = g_{k+1} - g_k$. The conjugacy condition (1.7) is effective since it leads to the efficient CG method proposed by Hestenes and Stiefel (HS) [11] in which the CG parameter, namely β_k^{HS}, is computed by

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}.$$

To find more efficient CG methods following the HS method, researchers paid special attention to modify the CG parameter β_k^{HS} by applying the function and the Hessian matrix information (see [12] and the references therein). For example, Perry [13] noted that the search direction d_{k+1} in the HS method can be written as

$$d_{k+1} = -\left(I - \frac{d_k y_k^T}{d_k^T y_k}\right) g_{k+1} = -\left(I - \frac{s_k y_k^T}{s_k^T y_k}\right) g_{k+1}.$$

So, he proposed a modification on the HS method in which the search direction d_{k+1} is computed by

$$d_{k+1} = -\underbrace{\left(I - \frac{s_k y_k^T}{s_k^T y_k} + \frac{s_k s_k^T}{s_k^T y_k}\right)}_{P_{k+1}} g_{k+1} = -P_{k+1} g_{k+1}. \tag{1.8}$$

Perry justified the addition of the correction term $s_k s_k^T / s_k^T y_k$ by noting that the matrix P_{k+1} in (1.8) satisfies the following equation:

$$y_k^T P_{k+1} = s_k^T,$$

which is similar, but not identical, to the standard secant equation [3,4] defined by

$$\nabla^2 f(x_{k+1}) s_k = y_k, \quad \text{or equivalently, } \nabla^2 f(x_{k+1})^{-1} y_k = s_k. \tag{1.9}$$

To improve Perry's method, Shanno [14] modified the matrix P_{k+1} as follows:

$$P_{k+1}^S \overset{\text{def}}{=} I - \frac{s_k y_k^T + y_k s_k^T}{s_k^T y_k} + \left(1 + \frac{y_k^T y_k}{s_k^T y_k}\right) \frac{s_k s_k^T}{s_k^T y_k}. \tag{1.10}$$

Thus, the related CG method is precisely the BFGS method proposed by Broyden [15], Fletcher [16], Goldfarb [17] and Shanno [18], in which the approximation of the inverse Hessian is restarted as the identity matrix at every step and so, no significant storage is needed to develop a better approximation for the inverse Hessian (this idea was first discussed by Nazareth [19] and Buckley [20]). It can be shown that if the line search procedure guarantees that $s_k^T y_k > 0$, then P_{k+1}^S defined by (1.10) is a positive definite matrix [3,4] and consequently, the search directions generated by $d_{k+1} = -P_{k+1}^S g_{k+1}$, for all $k \geq 0$, are descent directions.

In a different effort to enhance the efficiency of the CG methods, Birgin and Martínez [21] proposed a spectral CG method in which the search directions are defined by

$$d_0 = -g_0, \qquad d_{k+1} = -\theta_{k+1} g_{k+1} + \beta_k d_k, \quad k = 0, 1, 2, \ldots, \tag{1.11}$$