European Journal of Operational Research 234 (2014) 625–630

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Continuous Optimization

The Dai–Liao nonlinear conjugate gradient method with optimal parameter choices

Saman Babaie-Kafaki [a,*], Reza Ghanbari [b]

[a] Department of Mathematics, Faculty of Mathematics, Statistics and Computer Sciences, Semnan University, P.O. Box 35195-363, Semnan, Iran
[b] Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, P.O. Box 9177948953, Mashhad, Iran

ARTICLE INFO

Article history:
Received 17 March 2013
Accepted 12 November 2013
Available online 23 November 2013

Keywords:
Nonlinear programming
Large-scale optimization
Conjugate gradient algorithm
Singular value
Global convergence

ABSTRACT

Minimizing two different upper bounds of the matrix which generates search directions of the nonlinear conjugate gradient method proposed by Dai and Liao, two modified conjugate gradient methods are proposed. Under proper conditions, it is briefly shown that the methods are globally convergent when the line search fulfills the strong Wolfe conditions. Numerical comparisons between the implementations of the proposed methods and the conjugate gradient methods proposed by Hager and Zhang, and Dai and Kou, are made on a set of unconstrained optimization test problems of the CUTEr collection. The results show the efficiency of the proposed methods in the sense of 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 (Dai et al., 1999) which made them popular for engineers and mathematicians engaged in solving large-scale problems in the following form:

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

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

$$x_0\in\mathbb{R}^n,\ x_{k+1}=x_k+s_k,\ s_k=\alpha_k d_k,\quad k=0,1,\ldots, \tag{1.1}$$

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

$$d_0=-g_0,\ d_{k+1}=-g_{k+1}+\beta_k d_k,\quad k=0,1,\ldots, \tag{1.2}$$

where $g_k=\nabla f(x_k)$ and β_k is a scalar called the CG (update) parameter.

The steplength α_k is usually chosen to satisfy certain line search conditions (Sun & Yuan, 2006). Among them, the so-called strong Wolfe conditions (Wolfe, 1969) have attracted special attention in the convergence analyses and the implementations of CG methods, requiring that

$$f(x_k+\alpha_k d_k)-f(x_k)\leqslant \delta\alpha_k\nabla f(x_k)^T d_k, \tag{1.3}$$

$$|\nabla f(x_k+\alpha_k d_k)^T d_k|\leqslant -\sigma\nabla f(x_k)^T d_k, \tag{1.4}$$

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

Different choices for the CG parameter lead to different CG methods (Hager & Zhang, 2006b). Based on an extended conjugacy condition, one of the essential CG methods has been proposed by Dai and Liao (2001) (DL), with the following CG parameter:

$$\beta_k^{DL}=\frac{g_{k+1}^T y_k}{d_k^T y_k}-t\frac{g_{k+1}^T s_k}{d_k^T y_k}, \tag{1.5}$$

where t is a nonnegative parameter and $y_k=g_{k+1}-g_k$. Note that if $t=0$, then β_k^{DL} reduces to the CG parameter proposed by Hestenes and Stiefel (1952). Also, the CG parameter proposed by Hager and Zhang (2005) (HZ), i.e.

$$\beta_k^{HZ}=\frac{g_{k+1}^T y_k}{d_k^T y_k}-2\frac{\|y_k\|^2}{d_k^T y_k}\frac{g_{k+1}^T d_k}{d_k^T y_k}, \tag{1.6}$$

can be viewed as an adaptive version of (1.5) corresponding to $t=2\frac{\|y_k\|^2}{s_k^T y_k}$, where $\|\cdot\|$ denotes the Euclidean norm. Similarly, the CG parameter suggested by Dai and Kou (2013) (DK), i.e.

$$\beta_k(\tau_k)=\frac{g_{k+1}^T y_k}{d_k^T y_k}-\left(\tau_k+\frac{\|y_k\|^2}{s_k^T y_k}-\frac{s_k^T y_k}{\|s_k\|^2}\right)\frac{g_{k+1}^T s_k}{d_k^T y_k}, \tag{1.7}$$

in which τ_k is a parameter corresponding to the scaling factor in the scaled memoryless BFGS method (Sun & Yuan, 2006), can be considered as another adaptive version of (1.5) corresponding to

* Corresponding author. Tel.: +98 231 3354160.
E-mail addresses: sbk@semnan.ac.ir (S. Babaie-Kafaki), rghanbari@um.ac.ir (R. Ghanbari).

0377-2217/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2013.11.012

$t = \tau_k + \frac{\|y_k\|^2}{s_k^T y_k} - \frac{s_k^T y_k}{\|s_k\|^2}$. In (Dai & Liao, 2001), it has been shown that a CG method in the form of (1.1) and (1.2) with $\beta_k = \beta_k^{DL}$ is globally convergent for uniformly convex functions.

The approach of Dai and Liao has been paid special attention to by many researches. In several efforts, modified secant equations have been applied to make modifications on the DL method. For example, Yabe and Takano (2004) used the modified secant equation proposed by Zhang, Deng, and Chen (1999). Also, Zhou and Zhang (2006) applied the modified secant equation proposed by Li and Fukushima (2001). Li, Tang, and Wei (2007) used the modified secant equation proposed by Wei, Li, and Qi (2006). Ford, Narushima, and Yabe (2008) employed the multi-step quasi-Newton equations proposed by Ford and Moghrabi (1994). Babaie-Kafaki, Ghanbari, and Mahdavi-Amiri (2010) applied a revised form of the modified secant equation proposed by Zhang et al. (1999) and the modified secant equation proposed by Yuan (1991). Furthermore, in several other attempts, the modified versions of β_k^{DL} suggested in (Babaie-Kafaki et al., 2010; Ford et al., 2008; Li et al., 2007; Yabe & Takano, 2004; Zhou & Zhang, 2006) have been used to achieve descent CG methods. Examples include the studies made by Narushima and Yabe (2012), Sugiki, Narushima, and Yabe (2012), and Livieris and Pintelas (2012).

Here, based on a singular value study on the DL method, two nonlinear CG methods are proposed. The remainder of this work is organized as follows. In Section 2, the methods are suggested and their global convergence analysis is discussed. In Section 3, they are numerically compared with the CG methods proposed by Hager and Zhang, and Dai and Kou, and comparative testing results are reported. Finally, conclusions are made in Section 4.

2. Two modified nonlinear conjugate gradient methods

Based on Perry's point of view (Perry, 1976), it is notable that from (1.2) and (1.5), search directions of the DL method can be written as:

$$d_{k+1} = -Q_{k+1} g_{k+1}, \quad k = 0, 1, \ldots, \tag{2.1}$$

where

$$Q_{k+1} = I - \frac{s_k y_k^T}{s_k^T y_k} + t \frac{s_k s_k^T}{s_k^T y_k}.$$

So, the DL method can be considered as a quasi-Newton method in which the inverse Hessian is approximated by the nonsymmetric matrix Q_{k+1}. Since Q_{k+1} presents a rank-two update, its determinant can be computed by (Sun & Yuan, 2006, chap. 1)

$$\det(Q_{k+1}) = t \frac{\|s_k\|^2}{s_k^T y_k}. \tag{2.2}$$

Hence, if $t > 0$ and the line search guarantees that $s_k^T y_k \neq 0$, then Q_{k+1} is nonsingular.

It is remarkable that numerical performance of the DL method is very dependent on the parameter t for which there is no any optimal choice (Andrei, 2011). Motivated by this, here, based on a singular value study, two upper bounds for the condition number of the matrix Q_{k+1} are obtained and then, two optimal values for the parameter t are suggested. Now, we briefly discuss the singular value decomposition (SVD). For a more detailed discussion, see (Watkins, 2002, chap. 4).

Theorem 2.1 (Watkins, 2002, chap. 4). *(Geometric SVD Theorem) Let $A \in \mathbb{R}^{n\times m}$ be a nonzero matrix with rank r. Then, $\mathbb{R}^m$ has an orthonormal basis $v_1, \ldots, v_m$, $\mathbb{R}^n$ has an orthonormal basis $u_1, \ldots, u_n$, and there exist $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ such that*

$$A v_i = \begin{cases} \sigma_i u_i, & i = 1,2,\ldots,r, \\ 0, & i = r+1,\ldots,m, \end{cases} \quad \text{and} \quad A^T u_i = \begin{cases} \sigma_i v_i, & i = 1,2,\ldots,r, \\ 0, & i = r+1,\ldots,n. \end{cases}$$

Definition 2.1. The scalars $\{\sigma_i\}_{i=1}^r$ introduced in Theorem 2.1 are called the singular values of A.

Based on Theorem 2.1, for an arbitrary matrix $A \in \mathbb{R}^{n\times m}$ with rank r it can be seen that

$$\|A\|_F^2 = \sigma_1^2 + \cdots + \sigma_r^2, \tag{2.3}$$

where $\|\cdot\|_F$ stands for the Frobenius norm. Also, if $r = m = n$, then

$$|\det(A)| = \sigma_1 \times \sigma_2 \times \cdots \times \sigma_n. \tag{2.4}$$

One essential factor which plays an important role in the sensitivity analysis of a numerical problem related to a matrix, is the matrix condition number. For an arbitrary nonsingular matrix A, the scalar $\kappa(A)$ defined by

$$\kappa(A) = \|A\|\|A^{-1}\|,$$

is called the condition number of A. The matrix A with a large condition number is called an ill-conditioned matrix since the computations related to this matrix are potentially very sensitive to changes in the other data. Here, the following theorem is needed.

Theorem 2.2 (Watkins, 2002, chap. 4). *Let $A \in \mathbb{R}^{n\times n}$ be a nonsingular matrix with the singular values $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n > 0$. Then*

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}. \tag{2.5}$$

Remark 2.1. The condition number $\kappa(A)$ computed by (2.5) is called the spectral condition number.

In our analysis, we need to find the singular values of the matrix Q_{k+1}. Hereafter, we assume that $s_k^T y_k > 0$, as guaranteed by the strong Wolfe conditions (1.3) and (1.4).

Firstly, note that since $s_k^T y_k \neq 0$, there exists a set of mutually orthonormal vectors $\{u_k^i\}_{i=1}^{n-2}$ such that

$$s_k^T u_k^i = y_k^T u_k^i = 0, \quad i = 1, 2, \ldots, n-2,$$

which leads to

$$Q_{k+1} u_k^i = Q_{k+1}^T u_k^i = u_k^i, \quad i = 1, 2, \ldots, n-2.$$

That is, Q_{k+1} has $n-2$ singular values equal to 1. Next, we find the two remaining singular values of Q_{k+1} namely σ_k^- and σ_k^+.

Since $\|Q_{k+1}\|_F^2 = \mathrm{tr}(Q_{k+1}^T Q_{k+1})$, from (2.3) we get

$${\sigma_k^-}^2 + {\sigma_k^+}^2 = t^2 \frac{\|s_k\|^4}{(s_k^T y_k)^2} + \frac{\|s_k\|^2 \|y_k\|^2}{(s_k^T y_k)^2}. \tag{2.6}$$

Also, from (2.2) and (2.4) we have

$$\sigma_k^- \sigma_k^+ = t \frac{\|s_k\|^2}{s_k^T y_k}. \tag{2.7}$$

Now, from (2.6) and (2.7), after some algebraic manipulations we obtain the singular values σ_k^- and σ_k^+ as follows:

$$\sigma_k^{\pm} = \frac{1}{2} \frac{\sqrt{\left(t\|s_k\|^2 + s_k^T y_k\right)^2 + \|s_k\|^2\|y_k\|^2 - (s_k^T y_k)^2}}{s_k^T y_k} \pm \frac{1}{2} \times \frac{\sqrt{\left(t\|s_k\|^2 - s_k^T y_k\right)^2 + \|s_k\|^2\|y_k\|^2 - (s_k^T y_k)^2}}{s_k^T y_k}. \tag{2.8}$$

The following lemmas are now immediate.