An extended three-term conjugate gradient method with sufficient descent property

S. Babaie-Kafaki and R. Ghanbari
AN EXTENDED THREE-TERM CONJUGATE GRADIENT METHOD WITH SUFFICIENT DESCENT PROPERTY

S. BABAIE-KAFAKI AND R. GHANBARI

Received 27 May, 2014

Abstract. An extension of the three-term conjugate gradient method proposed by Zhang et al. is suggested. Based on an eigenvalue analysis on the search direction matrix, it is shown that the method possesses the sufficient descent property, no matter whether the line search is exact or not as well as the objective function is convex or not. It is interesting that the method can be considered as a hybridization of the conjugate gradient methods proposed by Zhang et al., and Hestenes and Stiefel. Global convergence of the method is established for uniformly convex objective functions. Comparative numerical results demonstrating efficiency of the proposed method are reported.

2010 Mathematics Subject Classification: 65K05; 90C53; 49M37; 15A18

Keywords: unconstrained optimization, large-scale optimization, three-term conjugate gradient method, eigenvalue, sufficient descent condition, global convergence

1. INTRODUCTION

Conjugate gradient (CG) methods comprise a class of unconstrained optimization algorithms with the attractive features of low memory requirement, simple computation and strong global convergence, making them efficient for solving large-scale problems in the form of \( \min_{x \in \mathbb{R}^n} f(x) \), with the differentiable nonlinear function \( f : \mathbb{R}^n \to \mathbb{R} \). Iterations of the CG methods are given by

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

where \( \alpha_k \) is a steplength to be computed by a line search procedure and \( d_k \) is the search direction defined by

\[
d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_k d_k, \quad k = 0, 1, \ldots
\]

in which \( g_k = \nabla f(x_k) \) and \( \beta_k \) is a scalar called the CG (update) parameter. Different CG methods mainly correspond to different choices for the CG parameter \([14]\). Although the CG methods are equivalent in the linear case, i.e., when the objective

This research was supported by Research Councils of Semnan University and Ferdowsi University of Mashhad.

© 2015 Miskolc University Press
function is strictly convex and the steplength is computed exactly, their numerical behavior for general functions may be quite different [1].

One of the essential CG methods has been proposed by Hestenes and Stiefel [15] (HS), with the following CG parameter:

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

where $y_k = s_{k+1} - s_k$. Although the HS method is numerically efficient [1] and has the nice property of satisfying the conjugacy condition $d_k^T y_k = 0$, $\forall k \geq 0$, independent of the line search and the objective function convexity, the method lacks the following descent property:

$$g_k^T d_k < 0, \forall k \geq 0.$$  \hspace{1cm} (1.2)

This motivated the researchers to deal with modifications of the HS method with descent property (1.2), or its stronger version called the sufficient descent property, i.e.,

$$g_k^T d_k \leq -c ||g_k||^2, \forall k \geq 0,$$  \hspace{1cm} (1.3)

where $c$ is a positive constant and $||.||$ stands for the Euclidean norm. In this context, Hager and Zhang [12] (HZ) proposed the following CG parameter:

$$\beta_k^{HZ} = \beta_k^{HS} - 2 \frac{||y_k||^2}{d_k^T y_k} \frac{g_{k+1}^T d_k}{d_k^T y_k},$$  \hspace{1cm} (1.4)

which guarantees (1.3) with $c = 7/8$. Afterwards, the following extension of $\beta_k^{HZ}$ has been proposed in [14]:

$$\beta_k^\theta = \beta_k^{HS} - \theta_k \frac{||y_k||^2}{d_k^T y_k} \frac{g_{k+1}^T d_k}{d_k^T y_k},$$  \hspace{1cm} (1.5)

where $\theta_k$ is a nonnegative parameter. It is important that if $\theta_k \geq \tilde{\theta} > 1/4$, then the sufficient descent condition (1.3) is achieved with $c = 1 - 1/(4\tilde{\theta})$ [2]. Recently, Dai and Kou [8] (DK) proposed the following family of CG parameters:

$$\beta_k^{DK} = \beta_k^{HS} - \left( \tau_k + \frac{||y_k||^2}{s_k^T y_k} - \frac{s_k^T y_k}{||s_k||^2} \right) \frac{g_{k+1}^T s_k}{d_k^T y_k},$$  \hspace{1cm} (1.6)

and established the sufficient descent property (1.3) for four special choices of $\tau_k$ which is a nonnegative parameter corresponding to the scaling factor in the scaled memoryless BFGS method [18]. More recently, Babaie-Kafaki and Ghanbari [3] proposed the following family of CG parameters:

$$\beta_k^{p,q} = \beta_k^{HS} - \left( \frac{p}{s_k^T y_k} - \frac{q}{||y_k||^2} \right) \frac{g_{k+1}^T s_k}{d_k^T y_k}.$$  \hspace{1cm} (1.7)
with the real parameters $p$ and $q$, guaranteeing the descent property (1.2) for $p > 1/4$ and $q \leq 1/4$. Note that all the CG parameters $\beta_k^{HZ}$ defined by (1.4), $\beta_k^q$ defined by (1.5), $\beta_k^{DK}$ defined by (1.6), and $\beta_k^{p,q}$ defined by (1.7) can be viewed as adaptive versions of the CG parameter proposed by Dai and Liao [9], i.e.,

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

where $t$ is a nonnegative parameter to be chosen appropriately. In another attempt, Zhang et al. [21] (ZZL) proposed a computationally efficient three-term CG method with the following search directions:

$$d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_k^{HS} d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k, \quad \forall k \geq 0. \tag{1.8}$$

which satisfies the sufficient descent condition $g_k^T d_k = -||g_k||^2, \quad \forall k \geq 0$. It is interesting that under the exact line search, since we have $g_{k+1}^T d_k = 0$, all the CG methods of HZ, DK, ZZL and the descent CG method with the parameter (1.7) reduce to the HS method; hence, they can be considered as descent extensions of the HS method.

Several other attempts have been made to develop the three-term CG methods with the important sufficient descent property. Examples include the CG methods proposed by Cheng [6], Cheng and Liu [7], Narushima et al. [16], Sugiki et al. [17], Zhang et al. [19, 20], and Babaie-Kafaki and Ghanbari [4]. Here, we suggest an extension of the three-term CG method proposed by Zhang et al. [21]. Our method can be considered as a hybridization of the HS and ZZL methods. The remainder of this work is organized as follows. In Section 2, we propose our method and discuss its global convergence. In Section 3, we numerically compare our method with the CG Descent [13] and ZZL methods, and report comparative testing results. Finally, we make conclusions in Section 4.

2. A THREE-TERM CONJUGATE GRADIENT METHOD

Here, we deal with an extended form of the search directions (1.8), obtained by embedding a parameter $t_k \in [0, 1]$ on the third term of $d_{k+1}, \quad k \geq 0$. More exactly, the search directions of our method, namely the EZZL method, are given by

$$d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_k^{HS} d_k - t_k \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k, \quad \forall k \geq 0. \tag{2.1}$$

Note that if $t_k = 0, \forall k \geq 0$, or the exact line search is used, then the method reduces to the HS method. Also, if $t_k = 1, \forall k \geq 0$, then the method is precisely the ZZL method. Hence, the three-term CG method EZZL can be considered as a hybridization of the HS and ZZL methods with the hybridization parameter $t_k$. Next, we compute the
parameter $t_k$ such that the EZZL method satisfy the sufficient descent condition (1.3). In what follows, we assume that $d_k^T y_k \neq 0$, $\forall k \geq 0$, as guaranteed by the popular Wolfe conditions [18], i.e.,

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

(2.2)

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f(x_k)^T d_k,$$

(2.3)

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

It is worth noting that from (2.1) the search directions of the EZZL method can be written as:

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

(2.4)

where

$$Q_{k+1} = I - \frac{s_k y_k^T}{s_k^T y_k} + t_k \frac{y_k s_k^T}{s_k^T y_k}, \quad t_k \in [0, 1].$$

Since from (2.4) we can write

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T Q_{k+1} g_{k+1} = -g_{k+1}^T \frac{Q_{k+1}^T + Q_{k+1}}{2} g_{k+1},$$

(2.5)

to study the sufficient descent property of the EZZL method with a proper value for $t_k$, in our analysis we need to find the eigenvalues of the following symmetric matrix:

$$A_{k+1} = \frac{Q_{k+1}^T + Q_{k+1}}{2} = I + \frac{t_k - 1}{2} \frac{s_k y_k^T}{s_k^T y_k} + t_k \frac{y_k s_k^T}{s_k^T y_k}.$$

(2.6)

Because $\alpha_k > 0$ and $d_k^T y_k \neq 0$, we have $s_k^T y_k \neq 0$ and consequently, $s_k \neq 0$ and $y_k \neq 0$. So, there exists a set of mutually orthonormal vectors $\{u'_i\}_{i=1}^{n-2}$ such that

$$s_k^T u'_i = y_k^T u'_i = 0, \quad i = 1, \ldots, n-2,$$

which leads to

$$A_{k+1} u'_i = u'_i, \quad i = 1, \ldots, n-2.$$

That is, the vectors $u'_i, i = 1, \ldots, n-2$, are the eigenvectors of $A_{k+1}$ corresponding to the eigenvalue 1. Next, we find the two remaining eigenvalues of $A_{k+1}$, namely $\lambda^-_k$ and $\lambda^+_k$.

Since the trace of a square matrix is equal to the sum of its eigenvalues, we have

$$\text{tr}(A_{k+1}) = n + t_k - 1 = \underbrace{1 + \ldots + 1}_{(n-2) \text{ times}} + \lambda^-_k + \lambda^+_k,$$

which leads to

$$\lambda^-_k + \lambda^+_k = t_k + 1.$$  

(2.7)

On the other hand, from the properties of the Frobenius norm, we have

$$||A_{k+1}||_F^2 = \text{tr}(A_{k+1}^T A_{k+1}) = \text{tr}(A_{k+1}^2)$$
AN EXTENDED THREE-TERM CONJUGATE GRADIENT METHOD

\[ \lambda_k^{-2} + \lambda_k^+ = \frac{1}{2} (t_k + 1)^2 + \frac{1}{2} (t_k - 1)^2 \frac{||s_k||^2 ||y_k||^2}{(s_k^T y_k)^2} \]

which leads to

\[ \lambda_k^{-2} + \lambda_k^+ = \frac{1}{2} (t_k + 1)^2 + \frac{1}{2} (t_k - 1)^2 \frac{||s_k||^2 ||y_k||^2}{(s_k^T y_k)^2}. \tag{2.8} \]

Now, from (2.7) and (2.8), after some algebraic manipulations we get

\[ \lambda_k^{-1} \lambda_k^+ = \frac{1}{4} (t_k + 1)^2 - \frac{1}{4} (t_k - 1)^2 \frac{||s_k||^2 ||y_k||^2}{(s_k^T y_k)^2}. \tag{2.9} \]

and consequently, from (2.7) and (2.9), \( \lambda_k^- \) and \( \lambda_k^+ \) can be computed by

\[ \lambda_k^\pm = \frac{1}{2} (1 + t_k) \pm \frac{1}{4} (1 - t_k) \frac{||s_k|| ||y_k||}{s_k^T y_k}. \]

From Cauchy-Schwarz inequality, it can be shown that \( \lambda_k^+ \geq 1 \geq \lambda_k^- \). So, in order to guarantee the positive definiteness of the matrix \( A_{k+1} \) defined by (2.6), it is necessary to have \( \lambda_k^\pm = \xi \in (0, 1] \). Thus, the parameter \( t_k \) should be computed by

\[ t_k = \frac{(2\xi - 1)(s_k^T y_k) + ||s_k|| ||y_k||}{s_k^T y_k + ||s_k|| ||y_k||}, \quad \xi \in (0, 1]. \tag{2.10} \]

It can be seen that for the parameter \( t_k \) defined by (2.10) we have \( t_k \in (0, 1] \). Also, since the matrix \( A_{k+1} \) with the parameter \( t_k \) given by (2.10) is positive definite and its smallest eigenvalue is equal to \( \xi \), from (2.5) we have

\[ d_{k+1}^T g_{k+1} = -(s_k^T A_{k+1} g_{k+1} \leq -\xi ||g_{k+1}||^2. \]

So, the following theorem is now immediate.

**Theorem 1.** The search directions defined by (2.1) in which the parameter \( t_k \) is computed by (2.10) satisfy the sufficient descent condition (1.3) with some positive constant \( c \in (0, 1] \).

Now, we discuss global convergence of the EZZL method. In our analysis, we need to make the following basic assumptions on the objective function.

**Assumption 1.** The level set \( \mathcal{L} = \{ x | f(x) \leq f(x_0) \} \), with \( x_0 \) to be the starting point of the iterative method (1.1), is bounded. Also, in some open convex neighborhood \( \mathcal{N} \) of \( \mathcal{L} \), \( f \) is continuously differentiable and its gradient is Lipschitz continuous; that is, there exists a positive constant \( L \) such that

\[ ||\nabla f(x) - \nabla f(y)|| \leq L ||x - y||, \quad \forall x, y \in \mathcal{N}. \tag{2.11} \]
Note that these assumptions imply that there exists a positive constant $\gamma$ such that
$$\|\nabla f(x)\| \leq \gamma, \ \forall x \in \mathcal{L} \quad \text{(see Proposition 3.1 of [5]).} \quad (2.12)$$

The following important lemma plays an essential role in proving the global convergence theorem of the EZZL method.

**Lemma 1 ([17]).** Suppose that Assumption 1 holds. Consider any iterative method in the form of (1.1) for which the sufficient descent condition (1.3) holds and the steplength $\alpha_k$ satisfies the Wolfe conditions (2.2) and (2.3). If
$$\sum_{k \geq 0} \frac{1}{||d_k||^2} = \infty,$$
then the method converges in the sense that
$$\liminf_{k \to \infty} ||g_k|| = 0. \quad (2.13)$$

Now, we can prove the following global convergence theorem for the EZZL method.

**Theorem 2.** Suppose that Assumption 1 holds. Consider the iterative method (1.1) in which the search direction $d_k$ is computed by (2.1) with the parameter $t_k$ given by (2.10), and the steplength $\alpha_k$ is determined such that the Wolfe conditions (2.2) and (2.3) are satisfied. If the objective function $f$ is uniformly convex on $\mathcal{N}$, then the method converges in the sense that (2.13) holds.

**Proof.** At first, note that from Theorem 1, because the search directions of the method are descent directions, from (2.2) the sequence $\{x_k\}_{k \geq 0}$ is a subset of the level set $\mathcal{L}$. Also, $d_k \neq 0$, $\forall k \geq 0$, and consequently, using Lemma 1, to complete the proof it is enough to show that $||d_k||$ is bounded above.

The uniform convexity of the differentiable function $f$ ensures that there exists a positive constant $\mu$ such that
$$s_k^T y_k \geq \mu ||s_k||^2 \quad \text{(see Theorem 1.3.16 of [18]).} \quad (2.14)$$

Now, from (2.1), (2.11), (2.12), (2.14), and since for $t_k$ defined by (2.10) we have $t_k \in (0, 1]$, we get
$$||d_{k+1}|| \leq ||g_{k+1}|| + \frac{||y_k|| ||s_{k+1}||}{s_k^T y_k} ||s_k|| + t_k \frac{||s_{k+1}|| ||s_k||}{s_k^T y_k} ||y_k|| \leq \gamma + \frac{L ||s_k|| \gamma}{\mu ||s_k||^2 ||s_k||} + \frac{\gamma ||s_k||}{\mu ||s_k||^2 ||s_k||} L ||s_k|| L ||s_k|| = \left(1 + 2 \frac{L}{\mu}\right) \gamma,$$
which completes the proof. \qed
3. Numerical experiments

Here, we present some numerical results obtained by applying C++ implementations of our extended three-term CG method EZZL with the search directions (2.1) in which the parameter $t_k$ is computed by (2.10), the three-term CG method ZZL proposed by Zhang et al. [21] with the search directions (1.8), and the CG Descent algorithm [13]. The codes were run on a PC with 3.2 GHz Intel I3 of CPU, 4 GB of RAM and Centos 6.2 server Linux operation system. Furthermore, the experiments were performed on a set of 145 unconstrained optimization test problems of the CUTEr collection [11], with default dimensions as given in Hager’s home page: ‘http://www.math.ufl.edu/~hager/’.

In our experiments, we used CG Descent version 5.1. For the EZZL method, we set $\xi = 0.96$ in (2.10) because of the promising numerical results obtained among the different values $\xi \in \{0.02k\}_{k=1}^{50}$. Similar to the CG Descent algorithm, in the methods of EZZL and ZZL we used the effective approximate Wolfe conditions proposed by Hager and Zhang [13] in the line search procedure, with the same parameter values as suggested in [13]. Moreover, all attempts to solve the test problems were terminated when $\|g_k\|_\infty < 10^{-6}$.

Efficiency comparisons were made using the performance profile of Dolan and Moré [10], on the running time and the total number of function and gradient evaluations being equal to $N_f + 3N_g$ [13], where $N_f$ and $N_g$ respectively denote the number of function and gradient evaluations. Performance profile gives, for every $\omega \geq 1$, the proportion $p(\omega)$ of the test problems that each considered algorithmic variant has a performance within a factor of $\omega$ of the best. Figures 1 and 2 show the comparisons results. As the figures demonstrate, the EZZL method is preferable to the ZZL method while the CG Descent method outperforms the EZZL and ZZL methods, both in the perspectives of the total number of function and gradient evaluations, and the CPU time.

Since the CG methods have been mainly designed for solving large-scale problems, we made further comparisons of the three methods EZZL, ZZL and CG Descent, on the numerical results of the 64 out of 145 test problems with the minimum dimensions being equal to 1000. Figures 3 and 4 demonstrate the comparisons results. As the figures show, with respect to the total number of function and gradient evaluations, and also with respect to the running time, the EZZL method is preferable to the methods of CG Descent and ZZL. Thus, the EZZL method seem to be practically promising, especially for solving large-scale problems.

4. Conclusions

An extended three-term conjugate gradient method which can be considered as a hybridization of the three-term conjugate gradient method suggested by Zhang et al. [21] and the conjugate gradient method suggested by Hestenes and Stiefel [15] has...
Figure 1. Total number of function and gradient evaluations performance profiles for EZZL, ZZL and CG_Descent.

Figure 2. CPU time performance profiles for EZZL, ZZL and CG_Descent.
Figure 3. Total number of function and gradient evaluations performance profiles for EZZL, ZZL and CG_Descent in large-scale cases.

Figure 4. CPU time performance profiles for EZZL, ZZL and CG_Descent in large-scale cases.
been proposed. Based on an eigenvalue study, it has been shown that the method possesses the important sufficient descent property, independent of the line search and the objective function convexity. Global convergence of the method has been established for uniformly convex objective functions. Numerical experiments showed the efficiency of the proposed method in the sense of the performance profile introduced by Dolan and Moré [10], particularly for solving large-scale problems.

ACKNOWLEDGEMENT

The authors are grateful to Professor William W. Hager for providing the line search code. They also thank the anonymous referee for his/her useful comments helped to improve the presentation.

REFERENCES

[17] K. Sugiki, Y. Narushima, and H. Yabe, “Globally convergent three-term conjugate gradient meth-
ods that use secant conditions and generate descent search directions for unconstrained optimiza-

Authors’ addresses

S. Babaie-Kafaki
Semnan University, Department of Mathematics, Faculty of Mathematics, Statistics and Computer
Science, P.O. Box: 35195-363, Semnan, Iran
E-mail address: sbk@semnan.ac.ir

R. Ghanbari
Ferdowsi University of Mashhad, Faculty of Mathematical Sciences, P.O. Box: 9177948953, Mash-
had, Iran
E-mail address: rghanbari@um.ac.ir