Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL
Pacific Journal of Optimization
Vol. 9, No. 1, pp. 23–28, January 2013

A NEW PROOF FOR THE SUFFICIENT DESCENT CONDITION OF ANDREI'S SCALED CONJUGATE GRADIENT ALGORITHMS*

Saman Babaie-Kafaki

Abstract: Based on an eigenvalue study, we show that the search directions of Andrei's scaled nonlinear conjugate gradient algorithms satisfy the important sufficient descent condition.

Key words: *unconstrained optimization, large-scale optimization, conjugate gradient algorithm, BFGS update, eigenvalue, sufficient descent condition*

Mathematics Subject Classification: *65K05, 49M37, 90C53, 15A18*

1 Introduction

Conjugate gradient (CG) methods comprise a class of unconstrained optimization algorithms characterized by low memory requirements and strong global convergence properties [8]. Although CG methods are not the fastest or most robust algorithms 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. A nice review of different CG methods has been presented in [12].

Recently, by hybridizing the memoryless BFGS CG method suggested by Shanno [14] and the spectral CG method suggested by Birgin and Martínez [7], Andrei [1, 2, 3, 4, 5] proposed several scaled nonlinear CG algorithms in the following form,

$$\begin{aligned} &x_0\in\mathbb{R}^n,\\ &x_{k+1}=x_k+s_k,\ s_k=\alpha_k d_k,\ k=0,1,..., \end{aligned} \tag{1.2}$$

where d_k is the search direction defined by

$$\begin{aligned} &d_0=-g_0,\\ &d_{k+1}=-Q^*_{k+1}g_{k+1},\ k=0,1,..., \end{aligned} \tag{1.3}$$

*This research was in part supported by a grant from IPM (No. 90900023), and in part by the Research Council of Semnan University.

with $g_k = \nabla f(x_k)$ and the following matrix $Q^*_{k+1} \in \mathbb{R}^{n\times n}$,

$$Q^*_{k+1} = \theta_{k+1} I - \theta_{k+1} \frac{y_k s_k^T + s_k y_k^T}{s_k^T y_k} + \left(1 + \theta_{k+1} \frac{y_k^T y_k}{s_k^T y_k}\right) \frac{s_k s_k^T}{s_k^T y_k}, \tag{1.4}$$

in which $y_k = g_{k+1} - g_k$, and the scalar parameter θ_{k+1} is determined based on the self-scaling technique of Oren and Spedicato [13], that is,

$$\theta_{k+1} = \frac{s_k^T s_k}{s_k^T y_k}. \tag{1.5}$$

Also, α_k in (1.2) is a steplength to be computed by the Wolfe line search conditions [15], i.e.,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k \nabla f(x_k)^T d_k, \tag{1.6}$$
$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f(x_k)^T d_k, \tag{1.7}$$

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

In [2, 3, 4, 5], it has been shown that for the search direction d_{k+1} defined by (1.3) we have

$$g_{k+1}^T d_{k+1} \leq -\frac{(g_{k+1}^T s_k)^2}{s_k^T y_k}. \tag{1.8}$$

Since from (1.7) we can write

$$s_k^T y_k = s_k^T g_{k+1} - s_k^T g_k \geq -(1-\sigma) s_k^T g_k > 0, \tag{1.9}$$

inequality (1.8) ensures that the search directions (1.3) are descent directions. Also, Andrei's scaled nonlinear CG algorithms are numerically efficient. Specially, Andrei's accelerated scaled memoryless BFGS preconditioned CG algorithm [5] can be considered as one of the most efficient CG methods.

In a recent effort to make correction in the convergence analysis of the scaled CG algorithms proposed in [2, 3, 4, 5], Babaie-Kafaki [6] showed that the search directions (1.3) satisfy the effective sufficient descent condition, i.e.,

$$g_k^T d_k \leq -c||g_k||^2, \ \forall k \geq 0, \tag{1.10}$$

where c is a positive constant. Here, in order to present another proof by obtaining a lower bound for the eigenvalues of Q^*_{k+1}, we show that the search directions of Andrei's scaled nonlinear CG algorithms satisfy the sufficient descent condition (1.10).

2 On the Sufficient Descent Condition of the Scaled Conjugate Gradient Algorithms

Although the descent condition is often adequate [8], sufficient descent condition may be crucial in the convergence analysis of the CG methods [9, 11]. Also, satisfying in the sufficient descent condition is considered as a strength of a CG method in the literature [10, 12]. Here, to present our eigenvalue study on the sufficient descent condition of the search directions of Andrei's scaled CG algorithms, the following preliminaries are needed.

Definition 2.1 ([15]). A differentiable function f is said to be uniformly (or strongly) convex on a nonempty open convex set $\mathcal{S}$ if there exists a positive constant μ such that

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \mu ||x - y||^2, \ \forall x, y \in \mathcal{S}. \tag{2.1}$$

Assumption 2.2. The objective function f in (1.1) is continuously differentiable and its gradient is Lipschitz continuous on a neighborhood $\mathcal{N}$ of the level set $\mathcal{L}$ defined by

$$\mathcal{L} = \{x \in \mathbb{R}^n : \ f(x) \leq f(x_0)\}, \tag{2.2}$$

with x_0 to be the starting point of the iterative method (1.2)-(1.3); that is, there exists a positive constant L such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \ \forall x, y \in \mathcal{N}. \tag{2.3}$$

Suppose that the objective function f is uniformly convex on the neighborhood $\mathcal{N}$ of the level set $\mathcal{L}$ defined by (2.2). Since the search directions (1.3) are descent directions, from (1.6) the sequence $\{x_k\}_{k\geq 0}$ generated by the iterative method (1.2)-(1.3) is a subset of the level set $\mathcal{L}$ and consequently, from (2.1) we have

$$s_k^T y_k \geq \mu ||s_k||^2. \tag{2.4}$$

It is remarkable that the matrix Q_{k+1}^* is determined based on the BFGS update [15] in which the inverse Hessian is restarted as $\theta_{k+1} I$ in each iteration and so, no significant storage is needed to develop a better approximation for the inverse Hessian. Since from (1.9) the Wolfe conditions ensure that $s_k^T y_k > 0$, Q_{k+1}^* is positive definite [15] and consequently, it is nonsingular. Based on the relationship between the BFGS and DFP updates [15], it can be shown that the matrix $W_{k+1}^* \in \mathbb{R}^{n\times n}$ defined by

$$W_{k+1}^* = \frac{1}{\theta_{k+1}} I - \frac{1}{\theta_{k+1}} \frac{s_k s_k^T}{s_k^T s_k} + \frac{y_k y_k^T}{s_k^T y_k}, \tag{2.5}$$

is the inverse of Q_{k+1}^*. Hence, W_{k+1}^* is also a positive definite matrix.

Now, based on the above argument, we prove the following theorem, ensuring satisfaction of the sufficient descent condition (1.10) for the search directions (1.3).

Theorem 2.3. *Suppose that Assumption 2.2 holds for the objective function f in (1.1). For the iterative method (1.2)-(1.3), if f is uniformly convex on $\mathcal{N}$ and the steplength α_k is determined to fulfill the Wolfe line search conditions (1.6) and (1.7), then the search directions (1.3) satisfy the sufficient descent condition (1.10).*

Proof. Since Q_{k+1}^* and W_{k+1}^* are $n \times n$ symmetric positive definite matrices, they have n positive eigenvalues. To establish the theorem, at first we show that for all $k \geq 0$, the eigenvalues of Q_{k+1}^* are bounded below by a positive constant.

From (1.9) we have $s_k^T y_k > 0$ and consequently, $s_k \neq 0$ and $y_k \neq 0$. So, there exists a set of mutually orthogonal unit vectors $\{u_k^i\}_{i=1}^{n-2}$ such that

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

which leads to

$$Q_{k+1}^* u_k^i = \theta_{k+1} u_k^i, \ i = 1, ..., n-2.$$

Thus, the vectors u_k^i, $i = 1, ..., n-2$, are the eigenvectors of Q_{k+1}^* correspondent to the eigenvalue θ_{k+1}. Let $\lambda_{k,n-1}$ and $\lambda_{k,n}$ be the two remaining eigenvalues of Q_{k+1}^*. Since the trace of a square matrix is equal to the summation of its eigenvalues, from (1.4) and (2.5) we can write

$$\begin{aligned} \operatorname{tr}(Q_{k+1}^*) &= (n-2)\theta_{k+1} + \frac{s_k^T s_k}{s_k^T y_k}\left(1 + \theta_{k+1}\frac{y_k^T y_k}{s_k^T y_k}\right) \\ &= \underbrace{\theta_{k+1} + ... + \theta_{k+1}}_{(n-2) \text{ times}} + \lambda_{k,n-1} + \lambda_{k,n}, \end{aligned} \tag{2.6}$$

and,

$$\begin{aligned}\mathrm{tr}(W^*_{k+1}) &= \frac{n-1}{\theta_{k+1}} + \frac{y_k^T y_k}{s_k^T y_k} \\ &= \underbrace{\frac{1}{\theta_{k+1}} + \dots + \frac{1}{\theta_{k+1}}}_{(n-2)\text{ times}} + \frac{1}{\lambda_{k,n-1}} + \frac{1}{\lambda_{k,n}}.\end{aligned} \tag{2.7}$$

So, from (2.6) and (2.7) we get

$$\begin{aligned}\lambda_{k,n-1} + \lambda_{k,n} &= \theta_{k+1}\left(1 + \theta_{k+1}\frac{y_k^T y_k}{s_k^T y_k}\right), \\ \frac{1}{\lambda_{k,n-1}} + \frac{1}{\lambda_{k,n}} &= \frac{1}{\theta_{k+1}}\left(1 + \theta_{k+1}\frac{y_k^T y_k}{s_k^T y_k}\right),\end{aligned} \tag{2.8}$$

and as a results,

$$\lambda_{k,n-1}\lambda_{k,n} = \theta_{k+1}^2. \tag{2.9}$$

Assume that $\lambda_{k,n} \leq \lambda_{k,n-1}$. From (1.5), (2.3) and (2.4) we have

$$\frac{1}{L} \leq \theta_{k+1} \leq \frac{1}{\mu}, \tag{2.10}$$

which together with (2.3), (2.4), (2.8), (2.9) and (2.10) yields

$$\begin{aligned}\lambda_{k,n} = \theta_{k+1}^2 \frac{1}{\lambda_{k,n-1}} &\geq \theta_{k+1}^2 \frac{1}{\lambda_{k,n-1} + \lambda_{k,n}} \\ &= \frac{s_k^T s_k}{s_k^T y_k + \theta_{k+1} y_k^T y_k} \geq \frac{\mu}{L^2 + L\mu}.\end{aligned} \tag{2.11}$$

Now, from (1.3) and (2.11), for all $k \geq 0$ we have

$$d_{k+1}^T g_{k+1} = -g_{k+1} Q^*_{k+1} g_{k+1} \leq -\lambda_{k,n}||g_{k+1}||^2 \leq -\frac{\mu}{L^2 + L\mu}||g_{k+1}||^2. \tag{2.12}$$

So, to complete the proof it is enough to let $c = \dfrac{\mu}{L^2 + L\mu}$ in (1.10).

□

Theorem 2.3 is necessary to complete the convergence analysis of the scaled CG algorithms proposed in [2, 3, 4, 5]. A different proof for this theorem has been proposed in [6]. More precisely, in the proof of Theorem 1 of [6], the following sufficient descent condition for the Andrei's scaled CG algorithms has been established,

$$d_{k+1}^T g_{k+1} \leq -\bar{c}||g_{k+1}||^2, \ \bar{c} = \frac{\mu}{L^2 + (n-1)L\mu}, \ \forall k \geq 0. \tag{2.13}$$

Since if $n > 2$, then $\bar{c} < \dfrac{\mu}{L^2 + L\mu}$, the newly established sufficient descent condition (2.12) is stronger than the old one stated by (2.13).

Acknowledgements

The author is grateful to the anonymous reviewers for their valuable suggestions helped to improve the presentation.

References

[1] N. Andrei, A scaled BFGS preconditioned conjugate gradient algorithm for unconstrained optimization, *Appl. Math. Lett.* 20 (2007) 645–650.

[2] N. Andrei, Scaled conjugate gradient algorithms for unconstrained optimization, *Comput. Optim. Appl.* 38 (2007) 401–416.

[3] N. Andrei, Scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization, *Optim. Methods Softw.* 22 (2007) 561–571.

[4] N. Andrei, A scaled nonlinear conjugate gradient algorithm for unconstrained optimization, *Optimization* 57 (2008) 549–570.

[5] N. Andrei, Accelerated scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization, *European J. Oper. Res.* 204 (2010) 410–420.

[6] S. Babaie-Kafaki, A note on the global convergence theorem of the scaled conjugate gradient algorithms proposed by Andrei, *Comput. Optim. Appl.* 52 (2012) 409414.

[7] E. Birgin and J.M. Martínez, A spectral conjugate gradient method for unconstrained optimization, *Appl. Math. Optim.* 43 (2001) 117–128.

[8] Y.H. Dai, J.Y. Han, G.H. Liu, D.F. Sun, H.X. Yin and Y.X. Yuan, Convergence properties of nonlinear conjugate gradient methods, *SIAM J. Optim.* 10 (1999) 348–358.

[9] Y.H. Dai and L.Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Appl. Math. Optim.* 43 (2001) 87–101.

[10] Y.H. Dai and Y.X. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.* 10 (1999) 177–182.

[11] J.C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optim.* 2 (1992) 21–42.

[12] W.W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, *Pac. J. Optim.* 2 (2006) 35–58.

[13] S.S. Oren and E. Spedicato, Optimal conditioning of self-scaling variable metric algorithms, *Math. Program.* 10 (1976) 70–90.

[14] D.F. Shanno, Conjugate gradient methods with inexact searches, *Math. Oper. Res.* 3 (1978) 244–256.

[15] W. Sun and Y.X. Yuan, *Optimization Theory and Methods: Nonlinear Programming*, Springer, New York, 2006.

Manuscript received 25 February 2012
revised 9 May 2012
accepted for publication 10 May 2012

Saman Babaie-Kafaki
Department of Mathematics, Faculty of Mathematics
Statistics and Computer Sciences, Semnan University
P.O. Box: 35195-363, Semnan, Iran
School of Mathematics, Institute for Research in Fundamental Sciences (IPM)
P.O. Box: 19395-5746, Tehran, Iran
E-mail address: sbk@semnan.ac.ir