Bull. Iranian Math. Soc.
Vol. 40 (2014), No. 1, pp. 235–242
Online ISSN: 1735-8515

AN EIGENVALUE STUDY ON THE SUFFICIENT DESCENT PROPERTY OF A MODIFIED POLAK-RIBIÈRE-POLYAK CONJUGATE GRADIENT METHOD

S. BABAIE KAFAKI

(Communicated by Abbas Salemi)

Abstract. Based on an eigenvalue analysis, a new proof for the sufficient descent property of the modified Polak-Ribière-Polyak conjugate gradient method proposed by Yu et al. is presented.
Keywords: Unconstrained optimization, conjugate gradient algorithm, sufficient descent condition, eigenvalue.
MSC(2010): Primary: 65K05; Secondary: 90C53, 49M37, 15A18.

1. Introduction

Conjugate gradient (CG) methods comprise a class of unconstrained optimization algorithms characterized by low memory requirements and strong global convergence properties [7]. Although CG methods are not the fastest or most robust optimization algorithms available today, they remain very popular for engineers and mathematicians engaged in solving large-scale problems in the following form:

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

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

$$\begin{aligned}&x_0\in\mathbb{R}^n,\\&x_{k+1}=x_k+s_k,\ s_k=\alpha_k d_k,\ k=0,1,...,\end{aligned}\tag{1.1}$$

Article electronically published on February 25, 2014.
Received: 24 August 2012, Accepted: 12 January 2013.

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

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

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

Different choices for the CG parameter lead to different CG methods (see [11] and the references therein). One of the efficient CG methods has been proposed by Polak, Ribière [12] and Polyak [13] (PRP), with the following CG parameter:

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{||g_k||^2},$$

where $y_k = g_{k+1} - g_k$, and $||.||$ stands for the Euclidean norm. Numerical efficiency of the PRP method is related to an automatic restart feature which avoids jamming [14], i.e., the generation of many short steps without making significant progress to the minimum. More exactly, when the step s_k is small, the factor y_k in the numerator of β_k^{PRP} tends to zero. Therefore, β_k^{PRP} becomes small and the new search direction is approximately the steepest descent direction. Next, a brief discussion on the convergence results development of the PRP method will be provided. In this context, the following definition is needed.

Definition 1.1. *We say that the search direction d_k is a descent direction (or equivalently, satisfies the descent condition) iff*

$$g_k^T d_k < 0.$$

Also, we say that the search directions $\{d_k\}_{k \geq 0}$ satisfy the sufficient descent condition iff

$$g_k^T d_k \leq -\tau ||g_k||^2, \ \forall k \geq 0, \tag{1.3}$$

where τ is a positive constant.

In the past decades, efforts have been made to study the global convergence of the PRP method. For example, Polak and Ribière [12] showed that the PRP method is globally convergent when the objective function is uniformly convex and the line search is exact. Powell [15] established that for a general nonlinear function, if $||s_k||$ tends to zero, the line search is exact, and ∇f is Lipschitz continuous, then the PRP method is globally convergent. On the other hand, he constructed a three dimensional counter example and showed that the PRP method with the exact line search may cycle infinitely without convergence to a solution [15]. Hence, the assumption $||s_k|| \to 0$ is necessary in the Powell's convergence result. Yuan [19] proved that if the search directions of the PRP method are descent directions, then, under the Wolfe line search conditions [17], the method is globally convergent for uniformly convex functions. However, Dai [5] showed that even for the uniformly convex functions, the PRP method may fail to generate a descent direction. So, the PRP method lacks global convergence in certain circumstances.

In a recent effort to make a modification on the PRP method in order to achieve the sufficient descent condition, Yu et al. [18] proposed a modified form of β_k^{PRP} as follows:

$$\beta_k^{DPRP} = \beta_k^{PRP} - C\frac{||y_k||^2}{||g_k||^4} g_{k+1}^T d_k, \tag{1.4}$$

where C is a constant satisfying

$$C > \frac{1}{4}. \tag{1.5}$$

Note that if the line search is exact, then $g_{k+1}^T d_k = 0$, and consequently, the CG parameter β_k^{DPRP} reduces to β_k^{PRP}.

An interesting feature of the DPRP method is that its search directions satisfy the sufficient descent condition (1.3) with $\tau = 1 - \frac{1}{4C}$, independent of the line search and the objective function convexity. Also, the DPRP method is globally convergent under certain line search strategies [18]. Furthermore, numerical results of [18] showed that the DPRP method is numerically efficient.

2. On the sufficient descent property of the DPRP method

Although the descent condition is often adequate [7], sufficient descent condition may be crucial in the convergence analysis of the CG methods [1–4, 8, 9]. Also, satisfying in the sufficient descent condition is considered as a strength of a CG method [6, 10]. Here, based on an eigenvalue study, a new proof for the sufficient descent property of the DPRP method is presented. The proof nicely demonstrates the importance of the condition (1.5).

It is remarkable that from (1.2) and (1.4), the search directions of the DPRP method can be written as:

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

where $Q_{k+1} \in \mathbb{R}^{n\times n}$ is defined by

$$Q_{k+1} = I - \frac{d_k y_k^T}{||g_k||^2} + C\frac{||y_k||^2}{||g_k||^4} d_k d_k^T. \tag{2.2}$$

Since Q_{k+1} is determined based on a rank-two update, its determinant can be computed by

$$\det(Q_{k+1}) = C\frac{||y_k||^2||d_k||^2}{||g_k||^4} - \frac{d_k^T y_k}{||g_k||^2} + 1 \stackrel{\text{def}}{=} \xi_k. \tag{2.3}$$

(See equality (1.2.70) of [16].) The following proposition is now immediate.

Proposition 2.1. *If the inequality (1.5) holds, then*

$$4\xi_k > \frac{||y_k||^2||d_k||^2 - (d_k^T y_k)^2}{||g_k||^4} \stackrel{\text{def}}{=} \gamma_k. \tag{2.4}$$

Proof. Since $C > \frac{1}{4}$, we can write

$$\begin{aligned} \xi_k - \frac{1}{4}\gamma_k &= \left(C - \frac{1}{4}\right)\frac{||y_k||^2||d_k||^2}{||g_k||^4} + \frac{1}{4}\frac{(d_k^T y_k)^2}{||g_k||^4} - \frac{d_k^T y_k}{||g_k||^2} + 1 \\ &= \left(C - \frac{1}{4}\right)\frac{||y_k||^2||d_k||^2}{||g_k||^4} + \left(\frac{1}{2}\frac{d_k^T y_k}{||g_k||^2} - 1\right)^2 > 0, \end{aligned}$$

which completes the proof. □

From Cauchy inequality, γ_k defined in (2.4) is nonnegative and so, the inequality (2.1) ensures that $\xi_k > 0$. Thus, from (2.3) the matrix Q_{k+1} defined by (2.2) is nonsingular. The following theorem ensures the sufficient descent property of the DPRP method.

Theorem 2.2. *For a CG method in the form of (1.1)-(1.2) with the CG parameter β_k^{DPRP} defined by (1.4), the following sufficient descent condition holds:*

$$g_k^T d_k \leq -\left(1 - \frac{1}{4C}\right)||g_k||^2, \ \forall k \geq 0. \tag{2.5}$$

Proof. Since $d_0 = -g_0$, the sufficient descent condition (2.5) holds for $k = 0$. For the subsequent indices, from (2.1) we can write

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T Q_{k+1}^T g_{k+1} = -g_{k+1}^T \frac{Q_{k+1}^T + Q_{k+1}}{2} g_{k+1}. \tag{2.6}$$

So, in our proof we need to find the eigenvalues of the following symmetric matrix:

$$A_{k+1} = \frac{Q_{k+1}^T + Q_{k+1}}{2} = I + C\frac{||y_k||^2}{||g_k||^4} d_k d_k^T - \frac{1}{2}\frac{d_k y_k^T + y_k d_k^T}{||g_k||^2}.$$

Note that if $y_k = 0$, then $A_{k+1} = I$, and consequently, all the eigenvalues of A_{k+1} are equal to 1. Otherwise, since $d_k \neq 0$, there exists a set of mutually orthogonal vectors $\{u_k^i\}_{i=1}^{n-2}$ such that

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

which leads to

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

That is, the vectors u_k^i, $i = 1, ..., n-2$, are the eigenvectors of A_{k+1} correspondent to the eigenvalue 1. Now, we find the two remaining eigenvalues of A_{k+1} namely λ_k^- and λ_k^+.

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

$$\begin{aligned} \mathrm{tr}(A_{k+1}) &= n + \xi_k - 1 \\ &= \underbrace{1 + ... + 1}_{(n-2) \text{ times}} + \lambda_k^- + \lambda_k^+, \end{aligned}$$

which yields

$$\lambda_k^- + \lambda_k^+ = \xi_k + 1, \tag{2.7}$$

with ξ_k defined in (2.3). On the other hand, since from the properties of the Frobenius norm, we have

$$\begin{aligned} ||A_{k+1}||_F^2 &= \mathrm{tr}(A_{k+1}^T A_{k+1}) = \mathrm{tr}(A_{k+1}^2) \\ &= \underbrace{1 + ... + 1}_{(n-2) \text{ times}} + {\lambda_k^-}^2 + {\lambda_k^+}^2, \end{aligned}$$

after some algebraic manipulations we get

$$\begin{aligned} {\lambda_k^-}^2 + {\lambda_k^+}^2 &= \left(C\frac{||d_k||^2||y_k||^2}{||g_k||^4} + 1\right)^2 - 2C\frac{||d_k||^2||y_k||^2(d_k^T y_k)}{||g_k||^6} \\ &\quad + \frac{1}{2}\frac{||d_k||^2||y_k||^2 + (d_k^T y_k)^2}{||g_k||^4} - 2\frac{d_k^T y_k}{||g_k||^2} + 1, \end{aligned}$$

which together with (2.7) yields

$$\lambda_k^- \lambda_k^+ = \xi_k - \frac{1}{4}\gamma_k, \tag{2.8}$$

with γ_k defined in (2.4). Thus, from (2.7) and (2.8), λ_k^- and λ_k^+ can be computed as the solutions of the following quadratic equation:

$$\lambda^2 - (\xi_k + 1)\lambda + \xi_k - \frac{1}{4}\gamma_k = 0.$$

More precisely,

$$\lambda_k^{\pm} = \frac{\xi_k + 1 \pm \sqrt{(\xi_k - 1)^2 + \gamma_k}}{2}.$$

Since $C > \frac{1}{4}$, from Proposition 2.1 we have $\xi_k > \frac{1}{4}\gamma_k$, and consequently, $\lambda_k^- > 0$. Also, $\lambda_k^+ \geq \lambda_k^-$ and so, $\lambda_k^+ > 0$. Therefore, the symmetric matrix A_{k+1} is positive definite. Hence, considering (2.6), to complete the proof it is enough to show that $\lambda_k^- \geq 1 - \frac{1}{4C}$. In this context, we define the following function:

$$h(z) = \frac{z + 1 - \sqrt{(z - 1)^2 + \gamma_k}}{2},$$

which is nondecreasing. Since $C > \frac{1}{4}$, it can be shown that

$$\xi_k \geq C\gamma_k + 1 - \frac{1}{4C}.$$

Thus,

$$h(\xi_k) = \lambda_k^- \geq f\left(C\gamma_k + 1 - \frac{1}{4c}\right) = 1 - \frac{1}{4C},$$

which completes the proof.

□

Acknowledgments

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

References

[1] S. Babaie-Kafaki, Erratum to: Scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization, *Optim. Methods Softw.* **28** (2013), no. 1, 217–219..

[2] S. Babaie-Kafaki, A note on the global convergence theorem of the scaled conjugate gradient algorithms proposed by Andrei, *Comput. Optim. Appl.* **52** (2012), no. 2, 409–414.

[3] S. Babaie-Kafaki, On the sufficient descent property of the Shanno's conjugate gradient method, *Optim. Lett.* **7** (2013), no. 4, 831–837.

[4] S. Babaie-Kafaki, A new proof for the sufficient descent condition of Andrei's scaled conjugate gradient algorithms, *Pac. J. Optim.*, to apear.

[5] Y. H. Dai, Analyses of Conjugate Gradient Methods, Ph.D. Thesis, Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1997.

[6] Y. H. Dai, New properties of a nonlinear conjugate gradient method, *Numer. Math.* **89** (2001), no. 1, 83–98.

[7] 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** (2000) 348–358.

[8] Y. H. Dai and L. Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Appl. Math. Optim.* **43** (2001), no. 1, 87–101.

[9] J. C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optim.* **2** (1992), no. 1, 21–42.

[10] W. W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, *SIAM J. Optim.* **16** (2005), no. 1, 170–192.

[11] W. W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, *Pac. J. Optim.* **2** (2006), no. 1, 35–58.

[12] E. Polak and G. Ribière, Note sur la convergence de méthodes de directions conjuguées, *Rev. Française Informat. Recherche Opérationnelle* **3** (1969) 35–43.

[13] B. T. Polyak, The conjugate gradient method in extreme problems, *USSR Comp. Math. Math. Phys.* **9** (1969) 94–112.

[14] M. J. D. Powell, Restart procedures for the conjugate gradient method, *Math. Programming* **12** (1977), no. 2, 241–254.

[15] M. J. D. Powell, Nonconvex Minimization Calculations and the Conjugate Gradient Method, 122–141, Numerical Analysis (Dundee, 1983), Lecture Notes in Mathematics, 1066, Springer, Berlin, 1984.

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

[17] P. Wolfe, Convergence conditions for ascent methods, *SIAM Rev.* **11** (1969) 226–235.

[18] G. Yu, L. Guan and G. Li, Global convergence of modified Polak-Ribière-Polyak conjugate gradient methods with sufficient descent property, *J. Ind. Manag. Optim.* **4** (2008) 565–579.

[19] Y. X. Yuan, Analysis on the conjugate gradient method, *Optim. Methods Softw.* **2** (1993) 19–29.

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