Optim Lett (2009) 3:11–21
DOI 10.1007/s11590-008-0086-5

ORIGINAL PAPER

Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems

Gonglin Yuan

Received: 16 August 2007 / Accepted: 18 April 2008 / Published online: 9 May 2008

Abstract It is well known that the sufficient descent condition is very important to the global convergence of the nonlinear conjugate gradient method. In this paper, some modified conjugate gradient methods which possess this property are presented. The global convergence of these proposed methods with the weak Wolfe–Powell (WWP) line search rule is established for nonconvex function under suitable conditions. Numerical results are reported.

Keywords Unconstrained optimization · Line search · Conjugate gradient method · Sufficient descent

1 Introduction

The nonlinear conjugate gradient (CG) method plays a very important role for solving the unconstrained optimization problem

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

where $f : \Re^n \to \Re$ is continuously differentiable. The CG method is usually designed by the iterative form

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \ldots \tag{1.2}$$

This work is supported by Guangxi University SF grands X061041 and China NSF grands 10761001.

G. Yuan (✉)
College of Mathematics and Information Science, Guangxi University,
Nanning, Guangxi 530004, People's Republic of China
e-mail: glyuan@gxu.edu.cn

where x_k is the current iterate point, $\alpha_k > 0$ is a steplength, and d_k is the search direction defined by

$$d_k = \begin{cases} -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1 \\ -g_k, & \text{if } k = 0, \end{cases} \tag{1.3}$$

where g_k is the gradient $\nabla f(x_k)$ of $f(x)$ at the point x_k, and $\beta_k \in \Re$ which determines the different conjugate gradient methods [1,13,19,23] is a scalar. There are some well-known formulas which are given as follows

$$\beta_k^{\text{PRP}} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}, \ [15,16], \tag{1.4}$$

$$\beta_k^{\text{FR}} = \frac{g_k^T g_k}{\|g_{k-1}\|^2}, \ [7], \tag{1.5}$$

$$\beta_k^{\text{CD}} = \frac{g_k^T g_k}{-d_{k-1}^T g_{k-1}}, \ [6], \tag{1.6}$$

$$\beta_k^{\text{LS}} = \frac{g_k^T(g_k - g_{k-1})}{-d_{k-1}^T g_{k-1}}, \ [14], \tag{1.7}$$

$$\beta_k^{\text{HS}} = \frac{g_k^T(g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}, \ [12], \tag{1.8}$$

$$\beta_k^{\text{DY}} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}, \ [3], \tag{1.9}$$

where g_{k-1} and g_k are the gradients $\nabla f(x_{k-1})$ and $\nabla f(x_k)$ of $f(x)$ at the point x_{k-1} and x_k, respectively, $\|.\|$ denotes the Euclidian norm of vectors. Although these methods are equivalent [4,23] when f is a strictly convex quadratic function and α_k is calculated by the following exact minimization rule:

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k), \tag{1.10}$$

their behaviors for general objective functions may be far different. For general functions, Zoutendjk [25] proved the global convergence of FR method with exact line search. Although one would be satisfied with its global convergence, the FR method performs much worse than the PRP (HS, LS) method in real computations. Powell [18] analyzed a major numerical drawback of the FR method, namely, if a small step is generated away from the solution point, the subsequent steps may be also very short. On the other hand, in practical computation, the PRP method, the HS method, and the LS method are generally believed to be the most efficient conjugate gradient methods since these methods essentially perform a restart if a bad direction occurs. The convergences of the CD method, the DY method, and the FR method are established [3,6,7], however their numerical results are not so well.

Powell [17] gave a counter example which showed that there exist nonconvex functions on which the PRP method does not converge globally even if the exact line search is used. He suggested that β_k should not be less than zero, which is very important to the global convergence [4,18]. Considering this suggestion, Gilbert and Nocedal [8] proved that the modified PRP method $\beta_k^+ = \max\{0, \beta_k^{\mathrm{PRP}}\}$ with the WWP line search is globally convergent under the sufficient descent condition. Over the past few years, much effort has been put to find out new formulas for conjugate methods which have not only global convergence property for general functions but also good numerical performance [4,8]. In resent years, some good results on the nonlinear conjugate gradient method are given [2,9–11,13,20,22,24].

The sufficient descent condition

$$g_k^T d_k \le -c\|g_k\|^2, \quad \forall\, k \ge 0, \;\; c > 0 \text{ is a constant} \tag{1.11}$$

is very important, and it may be crucial for the global convergence of conjugate gradient methods [8]. Then it is an interesting task to design a conjugate gradient method which possesses this condition.

Motivated by the above observations, we present some new conjugate gradient methods which have sufficient descent condition and the property of the scalar $\beta_k \ge 0$. The global convergence of the new methods is established for nonconvex function under assumptions. Numerical results show that these methods are interesting.

In the next section, motivation and the new algorithm are given. The sufficient descent property and the global convergence of the proposed algorithm are proved in Sect. 3. The generalization of the new technique is proposed in Sect. 4. Numerical results and one conclusion are presented in Sect. 5 and in Sect. 6, respectively.

2 Algorithm

Yu [21] proposed some modified conjugate gradient formulas and got better results. Here we state his modified PRP formula as follows:

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

where $\mu > \frac{1}{4}$ is a constant, and $y_k = g_{k+1} - g_k$. Motivated by his idea and the discussion of the above section, we present our new PRP formula:

$$\beta_k^{\mathrm{MPRP}} = \beta_k^{\mathrm{PRP}} - \min\left\{\beta_k^{\mathrm{PRP}}, \frac{\mu\|y_k\|^2}{\|g_k\|^4} g_{k+1}^T d_k\right\}. \tag{2.1}$$

It is easy to see that this formula satisfies the conjugate condition and the scalar $\beta_k^{\mathrm{MPRP}} \ge 0$ is true. In the next section, we will show that our new formula has the sufficient descent condition (1.11) too.

Algorithm 1 (The modified PRP algorithm)

Step 0: Choose an initial point $x_0 \in \Re^n$, $\varepsilon \in (0, 1)$, $\mu > 0$. Set $d_0 = -g_0 = -\nabla f(x_0)$, $k := 0$.
Step 1: If $\|g_k\| \leq \varepsilon$, then stop; Otherwise go to the next step.
Step 2: Compute step size α_k by some line search rule.
Step 3: let $x_{k+1} = x_k + \alpha_k d_k$. If $\|g_{k+1}\| \leq \varepsilon$, then stop.
Step 4: Calculate the search direction

$$d_{k+1} = -g_{k+1} + \beta_k^{\mathrm{MPRP}} d_k. \tag{2.2}$$

Step 5: Set $k := k + 1$, and go to Step 2.

3 The sufficient descent property and the global convergence

The weak Wolfe–Powell (WWP) line search is designed to find a step length α_k satisfying

$$f(x_k + \alpha_k d_k) \leq f_k + \delta \alpha_k g_k^T d_k \tag{3.1}$$

and

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{3.2}$$

where $f_k = f(x_k)$, $\delta \in (0, 1/2)$, $\sigma \in (\delta, 1)$. Under this line search, the global convergence of the very effective PRP conjugate gradient method is still open and the sufficient descent condition (1.11) cannot be satisfied for the PRP method and the FR method. In this section, we will prove that Algorithm 1 is globally convergent under this line search rule. Now we give the sufficient descent property of our new formula without any line search.

Theorem 3.1 *Consider* (2.2), *then the condition* (1.11) *holds for all* $k \geq 0$.

Proof If $k = 0$, then $g_0^T d_0 = -\|g_0\|^2$, (1.11) holds. Now we prove that (1.11) holds for all $k \geq 0$. By (1.4), (2.1), and (2.2), we have

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_k^{\mathrm{MPRP}} d_k^T g_{k+1} \\ &= -\|g_{k+1}\|^2 + \left(\beta_k^{\mathrm{PRP}} - \min\left\{\beta_k^{\mathrm{PRP}}, \frac{\mu\|y_k\|^2}{\|g_k\|^4} g_{k+1}^T d_k\right\}\right) d_k^T g_{k+1} \\ &= -\|g_{k+1}\|^2 + \left(\frac{g_{k+1}^T y_k}{\|g_k\|^2} - \min\left\{\frac{g_{k+1}^T y_k}{\|g_k\|^2}, \frac{\mu\|y_k\|^2}{\|g_k\|^4} g_{k+1}^T d_k\right\}\right) d_k^T g_{k+1} \end{aligned} \tag{3.3}$$

Denote $u = \frac{\|g_k\|}{\sqrt{2\mu}} g_k$, $v = \frac{\sqrt{2\mu} g_{k+1}^T d_k}{\|g_k\|} y_k$. We discuss the above equation by the following two cases.

Case (i) $\beta_k^{\mathrm{PRP}} < \frac{\mu\|y_k\|^2}{\|g_k\|^4} g_{k+1}^T d_k$. By (3.3), we obtain $g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2$.

Case (ii) $\beta_k^{\mathrm{PRP}} \geq \frac{\mu\|y_k\|^2}{\|g_k\|^4} g_{k+1}^T d_k$. The Eq. (3.3) can be rewritten as

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \left(\frac{g_{k+1}^T y_k}{\|g_k\|^2} - \frac{\mu\|y_k\|^2}{\|g_k\|^4} g_{k+1}^T d_k \right) d_k^T g_{k+1} \\ &= \frac{d_k^T g_{k+1} g_{k+1}^T y_k - \|g_{k+1}\|^2 \|g_k\|^2 - \frac{\mu\|y_k\|^2}{\|g_k\|^2} (g_{k+1}^T d_k)^2}{\|g_k\|^2} \\ &= \frac{u^T v - \frac{1}{2}(\|u\|^2 + \|v\|^2)}{\|g_k\|^2} + \frac{-(1 - \frac{1}{4\mu})\|g_{k+1}\|^2 \|g_k\|^2}{\|g_k\|^2} \\ &\leq -(1 - \frac{1}{4\mu})\|g_{k+1}\|^2, \end{aligned}$$

where the last inequality follows the inequality $u^T v \leq \frac{1}{2}(\|u\|^2 + \|v\|^2)$. Let $c = \min\{1, 1 - \frac{1}{4\mu}\}$, then the conclusion of this theorem holds. The proof is complete.

In order to prove the convergence of the nonlinear conjugate gradient methods, the following assumptions are often needed [3,4,23].

Assumption 3.1 (i) The level set $\Omega = \{x \in \Re^n \mid f(x) \leq f(x_0)\}$ is bounded, where x_0 is a given point.

(ii) In an open convex set Ω_0 that contains Ω, f has a lower bound, is differentiable, and its gradient g is Lipschitz continuous, namely, there exists a constants $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall\, x, y \in \Omega_0. \tag{3.4}$$

Lemma 3.1 *Suppose that Assumption* 3.1 *holds. Let the sequence* $\{g_k\}$ *and* $\{d_k\}$ *be generated by Algorithm* 1*, and let the stepsize* α_k *be determined by the WWP line search* (3.1) *and* (3.2)*. Then the Zoutendijk condition* [25]

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty \tag{3.5}$$

holds.

Proof By Theorem 3.1 and (3.1), we have

$$f(x_{k+1}) \leq f_k + \delta \alpha_k g_k^T d_k \leq f_k \leq f_{k-1} \leq \cdots \leq f(x_0),$$

which implies that the sequence $\{f_k\}$ is bounded. Using Theorem 3.1 again, (3.2), and Assumption 3.1(ii), we obtain

$$-(1 - \sigma) g_k^T d_k \leq (g_{k+1} - g_k)^T d_k \leq \|g_{k+1} - g_k\| \|d_k\| \leq \alpha_k L \|d_k\|^2,$$

combining this with (3.1), we get

$$\frac{\delta(1-\sigma)}{L}\frac{(g_k^T d_k)^2}{\|d_k\|^2} \le f_k - f_{k+1}.$$

Since $\{f_k\}$ is bounded, we have

$$\frac{\delta(1-\sigma)}{L}\sum_{k=0}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2} \le (f_0 - f_1) + (f_1 - f_2) + \cdots = f_0 - f^{\infty} < +\infty.$$

This completes the proof.

Theorem 3.2 *Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 1 with the weak Wolfe–Powell line search. Suppose that there exists a positive constant α^* that satisfies $\alpha_k \ge \alpha^*$. Then*

$$\lim_{k\to\infty}\|g_k\| = 0 \tag{3.6}$$

holds.

Proof From Assumption 3.1(i), there exists a constant $M > 0$ such that

$$\|\alpha_k d_k\| = \|s_k\| \le M. \tag{3.7}$$

Combining (3.7) and $\alpha_k \ge \alpha^*$, we have

$$\|d_k\| \le \frac{M}{\alpha^*}. \tag{3.8}$$

By (3.8), (3.5), and (1.11), we obtain (3.6). The proof is complete.

4 Generalizations of the new technique

Observing the formula β_k^{MPRP} and β_k^{PRP}, we let the numerator be $Y_k = y_k = g_{k+1} - g_k$ and the denominator be $R_k = \|g_k\|^2$, then

$$\beta_k^{\text{MPRP}} = \beta_k^{\text{PRP}} - \min\left\{\beta_k^{\text{PRP}}, \frac{\mu\|y_k\|^2}{\|g_k\|^4} g_{k+1}^T d_k\right\} = \frac{g_{k+1}^T Y_k}{R_k} - \min\left\{\frac{g_{k+1}^T Y_k}{R_k}, \frac{\mu\|Y_k\|^2}{R_k^2} g_{k+1}^T d_k\right\}.$$

So we generalize the choices of $Y_k = g_{k+1}$ and $R_k = \|g_k\|^2$ to the FR formula: $\beta_k^{\text{FR}} = \frac{g_{k+1}^T g_{k+1}}{\|g_k\|^2} = \frac{g_{k+1}^T Y_k}{R_k}$, the modified FR formula is:

$$\beta_k^{\text{MFR}} = \frac{g_{k+1}^T Y_k}{R_k} - \min\left\{\frac{g_{k+1}^T Y_k}{R_k}, \frac{\mu\|Y_k\|^2}{R_k^2} g_{k+1}^T d_k\right\} = \beta_k^{\text{FR}} - \min\left\{\beta_k^{\text{FR}}, \frac{\mu\|g_{k+1}\|^2}{\|g_k\|^4} g_{k+1}^T d_k\right\}.$$

Motivated by the above generalization and the idea of Yu [21], similarly, we get other modified formulas as follows:

$$\beta_k^{\mathrm{MCD}} = \beta_k^{\mathrm{CD}} - \min\left\{\beta_k^{\mathrm{CD}}, \frac{\mu\|g_{k+1}\|^2}{(-d_k^T g_k)^2} g_{k+1}^T d_k\right\},$$

$$\beta_k^{\mathrm{MLS}} = \beta_k^{\mathrm{LS}} - \min\left\{\beta_k^{\mathrm{LS}}, \frac{\mu\|y_k\|^2}{(-d_k^T g_k)^2} g_{k+1}^T d_k\right\},$$

$$\beta_k^{\mathrm{MDY}} = \beta_k^{\mathrm{DY}} - \min\left\{\beta_k^{\mathrm{DY}}, \frac{\mu\|g_{k+1}\|^2}{(d_k^T y_k)^2} g_{k+1}^T d_k\right\},$$

$$\beta_k^{\mathrm{MHS}} = \beta_k^{\mathrm{HS}} - \min\left\{\beta_k^{\mathrm{HS}}, \frac{\mu\|y_k\|^2}{(d_k^T y_k)^2} g_{k+1}^T d_k\right\}.$$

Similar to the modified PRP formula (2.1), it is not difficult to prove that the above modified formulas also have the sufficient descent property (1.11) and the corresponding algorithms satisfy (3.6). Then we state them as follows, but omit the proof.

Theorem 4.1 *Let the scalar β_k of (1.3) be replaced by β_k^{MFR}, β_k^{MCD}, β_k^{MLS}, β_k^{MDY}, and β_k^{MHS}, respectively, then we have the sufficient descent property (1.11). Moreover, if β_k^{PRP} of (2.2) is replaced by the above formulas, respectively, the conditions in Theorem 3.2 hold. Then the corresponding algorithms satisfy (3.6) too.*

5 Numerical results

In this section, numerical experiments are reported. The unconstrained optimization problems with the given initial points can be found at:

www.ici.ro/camo/neculai/SCALCG/testuo.pdf,

which were collected by Neculai Andrei. We test the modified methods with the WWP line search and compare their performances with the performances of the normal methods. The stop criterions are given below: the program is stopped if the inequality $\|g(x_k)\| \le \varepsilon$ is satisfied or the inequality $\|g(x_k)\| \le \varepsilon(1 + |f(x_k)|)$ is satisfied, where $\varepsilon = 1.0D-5$.

All codes were written in Fortran and run on PC with 2.60GHz, CPU processor, 256MB memory, and Windows XP operating system. In the following figures, the parameters were chosen as $\delta = 1.0D-4$, $\sigma = 1.0D-1$, and $\mu = 0.5$. The detailed numerical results are listed on the web site

Http://blog.sina.com.cn/gonglinyuan.

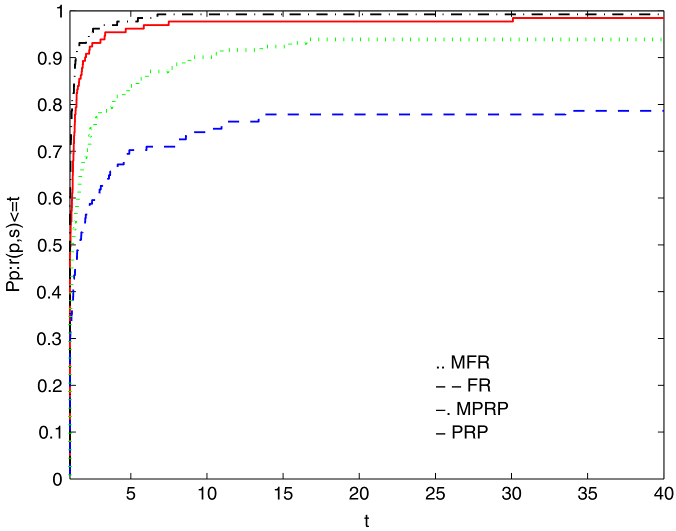

Fig. 1 Performance profiles of conjugate gradient methods MFR, FR, MPRP, and PRP

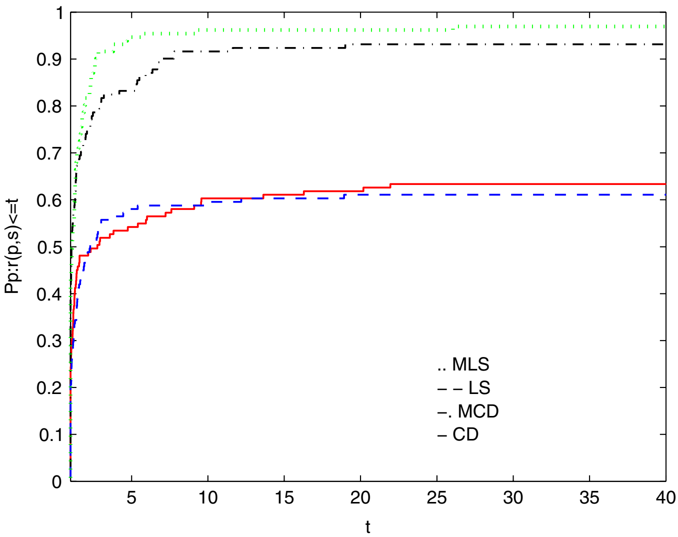

Fig. 2 Performance profiles of conjugate gradient methods MLS, LS, MCD, and CD

In Figs. 1–3, "PRP", "MPRP", "FR", "MFR", "CD", "MCD", "LS", "MLS", "DY", "MDY", "HS" and "MHS" stand for the PRP method, the modified PRP method, the FR method, the modified FR method, the CD method, the modified CD method, the LS method, the modified LS method, the DY method, the modified DY method, the HS method and the modified HS method, respectively. Figures 1–3 shows the performance of these twelve methods relative to NFN that denotes the sum of the iterative number

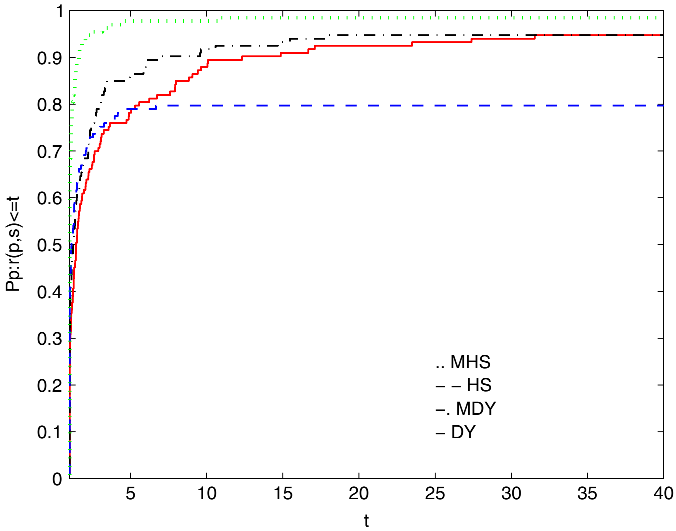

Fig. 3 Performance profiles of conjugate gradient methods MHS, HS, MDY, and DY

of the function value and the gradient value, which were evaluated using the profiles of Dolan and Moré [5].

In Fig. 1, we compare the performance of the MPRP method with the PRP method, the FR method, and the MFR method. Figure 1 shows that "MPRP" has best performance since it solves about 99% of the test problems successfully, and "PRP" has second best performance since it solves about 98% of the test problems successfully. "MFR" and "FR" can solve about 83 and 79% of the test problems successfully, respectively.

In Fig. 2, we compare the performance of the MCD method with the CD method, the LS method, and the MLS method. From Fig. 2, we see that "MLS" has the best performance and it can solve about 97% of the test problems successfully. The "MCD" outperforms "LS" and "CD" about 32 and 30%, respectively. Moreover, "MCD" and "CD" solve about 93 and 63% of the test problems, respectively. The "LS" performs worst since it cannot solve many problems. However, the "LS" is better than "CD" for $2 \leq t \leq 8$.

In Fig. 3, we compare the performance of the MHS method with the HS method, the DY method, and the MDY method. From Fig. 3, it is not difficult to see that the "MHS" performs best among these four methods and it can solve about 99% of the test problems successfully. The performance of "HS" is better than that of "DY" for $1 \leq t \leq 5$, and the "MDY" is better than "DY". The DY method, the HS method, and the MDY method solves about 95, 80, and 95% of the test problems successfully, respectively.

6 Conclusion

In this paper, we propose a modified PRP formula which possesses the sufficient descent condition without carrying out any line search. We generalize this technique to other conjugate gradient methods and get some modified conjugate gradient methods which have the sufficient descent property too. The global convergence of these methods is established under the weak Wolfe–Powell line search for nonconvex functions. Numerical results show that these given methods are competitive to the normal ones.

For further research, we should study the convergence of the new methods with other line search rules. Moreover, more numerical experiments for large practical problems and for the choice of the constant μ should be done in the future.

Acknowledgments We would like to thank Professor X. W. Lu and Z. X. Wei for their helpful conversations. We are also very grateful to anonymous referees and the editors for their valuable suggestions and comments, which improve our paper greatly.

References

1. Dai, Y.: A nonmonotone conjugate gradient algorithm for unconstrained optimization. J. Syst. Sci. Complex. **15**, 139–145 (2002)
2. Dai, Y., Liao, L.Z.: New conjugacy conditions and related nonlinear conjugate methods. Appl. Math. Optim. **43**, 87–101 (2001)
3. Dai, Y., Yuan, Y.: A nonlinear conjugate gradient with a strong global convergence properties. SIAM J. Optim. **10**, 177–182 (2000)
4. Dai, Y., Yuan, Y.: Nonlinear conjugate gradient Methods. Shanghai Scientific and Technical Publishers (1998)
5. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program. **91**, 201–213 (2002)
6. Fletcher, R.: Practical method of optimization, vol I: unconstrained optimization, 2nd edn. Wiley, New York (1997)
7. Fletcher, R., Reeves, C.: Function minimization bu conjugate gradients. Comput. J. **7**, 149–154 (1964)
8. Gibert, J.C., Nocedal, J.: Global convergence properties of conugate gradient methods for optimization. SIAM J. Optim. **2**, 21–42 (1992)
9. Hager, W.W., Zhang, H.: A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM J. Optim. **16**, 170–192 (2005)
10. Hager, W.W., Zhang, H.: Algorithm 851: $CG_DESCENT$, A conjugate gradient method with guaranteed descent. ACM Trans. Math. Softw. **32**, 113–137 (2006)
11. Hager, W.W., Zhang, H.: A survey of nonlinear conjugate gradient methods. Pac. J. Optim. **2**, 35–58 (2006)
12. Hestenes, M.R., Stiefel, E.: Method of conjugate gradient for solving linear equations. J, Res. Nat. Bur. Stand. **49**, 409–436 (1952)
13. Li, G., Tang, C., Wei, Z.: New conjugacy condition and related new conjugate gradient methods for unconstrained optimization problems. J. Comput. Appl. Math. **202**, 532–539 (2007)
14. Liu, Y., Storey, C.: Effcient generalized conjugate gradient algorithms part 1: theory. J. Comput. Appl. Math. **69**, 17–41 (1992)
15. Polak, E., Ribiere, G.: Note sur la xonvergence de directions conjugees. Rev. Francaise Informat Recherche Operatinelle, 3E Annee **16**, 35–43 (1969)
16. Polyak, B.T.: The conjugate gradient method in extreme problems. USSR Comp. Math. Math. Phys. **9**, 94–112 (1969)
17. Powell, M.J.D.: Nonconvex minimization calculations and the conjugate gradient method. Lecture Notes in Mathematics, vol. 1066, pp. 122–141. Spinger, Berlin (1984)
18. Powell, M.J.D.: Convergence properties of algorithm for nonlinear optimization. SIAM Rev. **28**, 487–500 (1986)
19. Wei, Z., Li, G., Qi, L.: New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems. Appl. Math. Comput. **179**, 407–430 (2006)
20. Wei, Z., Yao, S., Liu, L.: The convergence properties of some new conjugate gradient methods. Appl. Math. Comput. **183**, 1341–1350 (2006)

21. Yu, G.H.: Nonlinear self-scaling conjugate gradient methods for large-scale optimization problems. Thesis of Doctor's Degree, Sun Yat-Sen University (2007)
22. Yu, G., Guan, L.: Chen Spectral conjugate gradient methods with sufficient descent property for large-scale unconstrained optimization. Optimization methods and software (in press, 2007)
23. Yuan, Y., Sun, W.: Theory and methods of optimization. Science Press of China, Beijing (1999)
24. Zhang, L., Zhou, W., Li, D.: A descent modified Polak-Ribière-Polyak conjugate method and its global convergence. IMA J. Numer. Anal. **26**, 629–649 (2006)
25. Zoutendijk, G.: Nonlinear programming computational methods. In: Abadie J. (ed.) Integer and nonlinear programming, pp. 37–86. North-Holland, Amsterdam (1970)