Scientific
Research

A Line Search Algorithm for Unconstrained Optimization*

Gonglin Yuan[1], Sha Lu[2], Zengxin Wei[1]

[1]College of Mathematics and Information Science, Guangxi University, Nanning, China; [2]School of Mathematical Science, Guangxi Teachers Education University, Nanning, China.
Email: glyuan@gxu.edu.cn

ABSTRACT

It is well known that the line search methods play a very important role for optimization problems. In this paper a new line search method is proposed for solving unconstrained optimization. Under weak conditions, this method possesses global convergence and R-linear convergence for nonconvex function and convex function, respectively. Moreover, the given search direction has sufficiently descent property and belongs to a trust region without carrying out any line search rule. Numerical results show that the new method is effective.

Keywords: *Line Search, Unconstrained Optimization, Global Convergence, R-linear Convergence*

1. Introduction

Consider the unconstrained optimization problem

$$\min_{x \in R^n} f(x), \tag{1}$$

where $f : R^n \to R$ is continuously differentiable. The line search algorithm for (1) often generates a sequence of iterates $\{x_k\}$ by letting

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

where x_k is the current iterate point, d_k is a search direction, and $\alpha_k > 0$ is a steplength. Different choices of d_k and α_k will determine different line search methods [1-3]. The method is divided into two stages at each iteration: 1) choose a descent search direction d_k; 2) choose a step size α_k along the search direction d_k.

Throughout this paper, we denote $f(x_k)$ by f_k, $\nabla f(x_k)$ by g_k, and $\nabla f(x_{k+1})$ by g_{k+1}, respectively. $\| \cdot \|$ denotes the Euclidian norm of vectors.

One simple line search method is the steepest descent method if we take $d_k = -g_k$ as a search direction at every iteration, which has wide applications in solving

large-scale minimization problems [4]. However, the steepest descent method often yields zigzag phenomena in solving practical problems, which makes the algorithm converge to an optimal solution very slowly, or even fail to converge [5,6]. Then the steepest descent method is not the fastest one among the line search methods.

If $d_k = -H_k g_k$ is the search direction at each iteration in the algorithm, where H_k is an $n \times n$ matrix approximating $[\nabla^2 f(x_k)]^{-1}$, then the corresponding line search method is called Newton-like method [4-6] such as Newton method, quasi-Newton method, variable metric method, etc. Many papers [7-10] have been proposed by the method for optimization problems. However, one drawback of the Newton-like method is required to store and compute matrix H_k at each iteration and thus adds the cost of storage and computation. Accordingly, this method is not suitable to solve large-scale optimization problems in many cases.

The conjugate gradient method is a powerful line search method for solving the large scale optimization problems because of its simplicity and its very low memory requirement. The search direction of the conjugate gradient method often takes the form

$$d_k = \begin{cases} -g_k + \beta_k d_k, & if \ k \geq 1 \\ -g_k, & if \ k = 0, \end{cases} \tag{3}$$

where $\beta_k \in R$ is a scalar which determines the different

conjugate gradient methods [11-13] etc. The convergence behavior of the different conjugate gradient methods with some line search conditions [14] has been widely studied by many authors for many years (see [4, 15]). At present, one of the most efficient formula for β_k from the computation point of view is the following PRP method

$$\beta_k^{PRP} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{\| g_k \|^2}. \tag{4}$$

If $x_{k+1} \to x_k$, it is easy to get $\beta_k^{PRP} \to 0$, which implies that the direction d_k of the PRP method will turn out to be the steepest descent direction as the restart condition automatically when the next iteration point is approximate to the current point. This case is very important to ensure the efficiency of the PRP conjugate gradient method (see [4,15] etc.). For the convergence of the PRP conjugate gradient method, Polak and Ribière [16] proved that the PRP method with the exact line search is globally convergent when the objective function is convex, and Powell [17] gave a counter example to show that there exist nonconvex functions on which the PRP method does not converge globally even the exact line search is used.

We all know that the following sufficient descent condition

$$g_k^T d_k \le -c \| g_k \|^2, \text{ for all } k \ge 0 \text{ and some constant}$$

$$c > 0 \tag{5}$$

is very important to insure the global convergence of the algorithm by nonlinear conjugate gradient method, and it may be crucial for conjugate gradient methods [14]. It has been showed that the PRP method with the following strong Wolfe-Powell (SWP) line search rules which is to find the step size α_k satisfying

$$f(x_k + \alpha_k d_k) \le f_k + \sigma_1 \alpha_k g_k^T d_k, \tag{6}$$

and

$$| g(x_k + \alpha_k d_k)^T d_k | \le \sigma_2 | g_k^T d_k | \tag{7}$$

did not ensure the condition (5) at each iteration, where $0 < \sigma_1 < \frac{1}{2}$, $\sigma_1 < \sigma_2 < 1$. Then Grippo and Lucidi [18] presented a new line search rule which can ensure the sufficient descent condition and established the convergence of the PRP method with their line search technique. Powell [17] suggested that β_k should not be less than zero. Considering this idea, Gilbert and Nocedal [14] proved that the modified PRP method $\beta_k^+ = \max\{0, \beta_k^{PRP}\}$ is globally convergent under the sufficient descent assumption condition and the following weak Wolfe-Powell (WWP) line search technique: find the steplength α_k such that (6) and

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma_2 g_k^T d_k. \tag{8}$$

Over the past few years, much effort has been put to find out new formulas for conjugate methods such that they have not only global convergence property for general functions but also good numerical performance (see [14,15]). Resent years, some good results on the nonlinear conjugate gradient method are given [19-25].

These observations motivate us to propose a new method which possesses not only the simplicity and low memory but also desirable theoretical features. In this paper, we design a new line search method which possesses not only the sufficiently descent property but also the following property

$$\| d_k \| \le c_1 \| g_k \|, \text{ for all } k \ge 0 \text{ and some constant}$$

$$c_1 > 0 \tag{9}$$

whatever line search rule is used, where the property (9) implies that the search direction d_k is in a trust region automatically.

This paper is organized as follows. In the next section, the algorithms and other line search rules are stated. The global convergence and the R-linear convergence of the new method are established in Section 3. Numerical results and one conclusion are presented in Section 4 and in Section 5, respectively.

2. The Algorithms

Besides the inexact line search techniques WWP and SWP, there exist other line search rules which are often used to analyze the convergence of the line search method:

1) The exact minimization rule. The step size α_k is chosen such that

$$f(x_k + \alpha_k d_k) = \min_{\alpha \ge 0} f(x_k + \alpha d_k). \tag{10}$$

2) Goldstein rule. The step size α_k is chosen to satisfy (6) and

$$f(x_k + \alpha_k d_k) \ge f_k + \sigma_2 \alpha_k g_k^T d_k. \tag{11}$$

Now we give our algorithm as follows.
1) Algorithm 1 (New Algorithm)
Step 0: Choose an initial point $x_0 \in R^n$, and constants $0 < \varepsilon < 1$, $0 < \sigma_1 < \frac{1}{2}$, $\sigma_1 < \sigma_2 < 1$. Set $d_0 = -g_0 = -\nabla f(x_0)$, $k := 0$.

Step 1: If $\| g_k \| \le \varepsilon$, then stop; Otherwise go to step 2.

Step 2: Compute steplength α_k by one line search technique, and let $x_{k+1} = x_k + \alpha_k d_k$.

Step 3: If $\| g_{k+1} \| \le \varepsilon$, then stop; Otherwise go to step 4.

Step 4: Calculate the search direction d_{k+1} by (3), where β_k is defined by (4).

Step 5: Let $d_{k+1}^{new} = d_{k+1}' + \min\{0, -\frac{g_{k+1}^T d_{k+1}'}{\| g_{k+1} \|^2}\} g_{k+1} - g_{k+1}$,

where $d_{k+1}' = \frac{\| y_k^* \| \| g_{k+1} \|}{\| s_k \| \| d_{k+1} \|} d_{k+1}$, $s_k = x_{k+1} - x_k$,

$\| y_k^* \| = \max\{\| s_k \|, \| y_k \|\}$, $y_k = g_{k+1} - g_k$.

Step 6: Let $d := d_{k+1}^{naw}$, $k := k+1$, and go to step 2.

Remark. In the Step 5 of Algorithm 1, we have

$$\frac{\| y_k^* \|}{\| s_k \|} = \frac{\max\{\| s_k \|, \| y_k \|\}}{\| s_k \|} \geq 1,$$

which can increase the convergent speed of the algorithm from the computation point of view.

Here we give the normal PRP conjugate gradient algorithm and one modified PRP conjugate gradient algorithm [14] as follows.

2) Algorithm 2 (PRP Algorithm)

Step 0: Choose an initial point $x_0 \in R^n$, and constants $0 < \varepsilon < 1$, $0 < \sigma_1 < \frac{1}{2}$, $\sigma_1 < \sigma_2 < 1$. Set $d_0 = -g_0 = -\nabla f(x_0)$, $k := 0$.

Step 1: If $\| g_k \| \leq \varepsilon$, then stop; Otherwise go to step 2.

Step 2: Compute steplength α_k by one line search technique, and let $x_{k+1} = x_k + \alpha_k d_k$.

Step 3: If $\| g_{k+1} \| \leq \varepsilon$, then stop; Otherwise go to step 4.

Step 4: Calculate the search direction d_{k+1} by (3), where β_k is defined by (4).

Step 5: Let $k := k+1$ and go to step 2.

3) Algorithm 3 (PRP$^+$ Algorithm see [14])

Step 0: Choose an initial point $x_0 \in R^n$, and constants $0 < \varepsilon < 1$, $0 < \sigma_1 < \frac{1}{2}$, $\sigma_1 < \sigma_2 < 1$. Set $d_0 = -g_0 = -\nabla f(x_0)$, $k := 0$.

Step 1: If $\| g_k \| \leq \varepsilon$, then stop; Otherwise go to step 2.

Step 2: Compute steplength α_k by one line search technique, and let $x_{k+1} = x_k + \alpha_k d_k$.

Step 3: If $\| g_{k+1} \| \leq \varepsilon$, then stop; Otherwise go to step 4.

Step 4: Calculate the search direction d_{k+1} by (3), where $\beta_k = \max\{0, \beta_k^{PRP}\}$

Step 5: Let $k := k+1$ and go to step 2.

We will concentrate on the convergent results of Algorithm 1 in the following section.

3. Convergence Analysis

The following assumptions are often needed to analyze the convergence of the line search method (see [15,26]).

Assumption A (i) f is bounded below on the level set $\Phi = \{x \in R^n : f(x) \leq f(x_0)\}$;

Assumption A (ii) In some neighborhood Φ_0 of Φ, f is differentiable and its gradient is Lipschitz continuous, namely, there exists a constants $L > 0$ such that $\| g(x) - g(y) \| \leq L \| x - y \|$, for all $x, y \in \Phi_0$.

In the following, let $g_k \neq 0$ for all k, for otherwise a stationary point has been found.

Lemma 3.1 *Consider Algorithm* 1. *Let Assumption* (ii) *hold. Then* (5) *and* (9) *hold.*

Proof. If $k = 0$, (5) and (9) hold obviously. For $k \geq 1$, by Assumption (ii) and the Step 5 of Algorithm 1, we have

$$\| d_{k+1}^{new} \| \leq \| d_{k+1}' \| + \| d_{k+1}' \|$$
$$+ \| g_{k+1} \| \leq (2\max\{1, L\} + 1) \| g_{k+1} \|.$$

Now we consider the vector product $g_{k+1}^T d_{k+1}'$ in the following two cases:

case 1. If $g_{k+1}^T d_{k+1}' \leq 0$. Then we get

$$g_{k+1}^T d_{k+1}^{new} = g_{k+1}^T d_{k+1}' + \min\{0, -\frac{g_{k+1}^T d_{k+1}'}{\| g_{k+1} \|^2}\} \| g_{k+1} \|^2 - \| g_{k+1} \|^2$$
$$= g_{k+1}^T d_{k+1}' - \| g_{k+1} \|^2$$
$$\leq - \| g_{k+1} \|^2.$$

case 2. If $g_{k+1}^T d_{k+1}' > 0$. Then we obtain

$$g_{k+1}^T d_{k+1}^{new} = g_{k+1}^T d_{k+1}' + \min\{0, -\frac{g_{k+1}^T d_{k+1}'}{\| g_{k+1} \|^2}\} \| g_{k+1} \|^2 - \| g_{k+1} \|^2$$
$$= g_{k+1}^T d_{k+1}' - \frac{g_{k+1}^T d_{k+1}'}{\| g_{k+1} \|^2} \| g_{k+1} \|^2 - \| g_{k+1} \|^2$$
$$= - \| g_{k+1} \|^2.$$

Let $c \in (0,1)$, $c_1 = 2\max\{1, L\} + 1$ and use the Step 6 of Algorithm 1, (5) and (9) hold, respectively. The proof is completed.

The above lemma shows that the search direction d_k has such that the sufficient descent condition (5) and the condition (9) without any line search rule.

Based on Lemma 3.1, Assumption (i) and (ii), let us give the global convergence theorem of Algorithm 1.

Theorem 3.1 *Let* $\{\alpha_k, d_k, x_{k+1}, g_{k+1}\}$ *be generated by Algorithm* 1 *with the exact minimization rule, the Goldstein line search rule, the SWP line search rule, or the WWP line search rule, and Assumption* (i) *and* (ii) *hold. Then*

$$\lim_{k \to \infty} \| g_k \| = 0 \qquad (12)$$

holds.

Proof. We will prove the result of this theorem with the exact minimization rule, the Goldstein line search rule, the SWP line search rule, and the WWP line search rule, respectively.

1) For the exact minimization rule. Let the step size α_k be the solution of (10).

By the mean value theorem, $g_k^T d_k < 0$, and Assumption (ii), for any

$$\alpha_k^* \in \left[\frac{1}{5L} \frac{|g_k^T d_k|}{\|d_k\|^2}, \frac{2}{5L} \frac{|g_k^T d_k|}{\|d_k\|^2} \right],$$

we have

$$f(x_k + \alpha_k d_k) - f(x_k) \leq f(x_k + \alpha_k^* d_k) - f(x_k)$$
$$= \int_0^1 g(x_k + t\alpha_k^* d_k)^T (\alpha_k^* d_k) dt$$
$$= \alpha_k^* g_k^T d_k + \int_0^1 [g(x_k + t\alpha_k^* d_k) - g_k]^T (\alpha_k^* d_k) dt$$

$$\leq \alpha_k^* g_k^T d_k + \frac{1}{2} L \alpha_k^{*2} \|d_k\|^2$$

$$\leq -\frac{1}{5L} \frac{|g_k^T d_k|}{\|d_k\|^2}(-g_k^T d_k) + \frac{1}{2} L \frac{4}{25L^2} \frac{(g_k^T d_k)^2}{\|d_k\|^4} \|d\|^2 \quad (13)$$

$$= -\frac{3}{25L} \frac{(g_k^T d_k)^2}{\|d_k\|^2},$$

which together with Assumption (i), we can obtain

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty . \quad (14)$$

This implies that

$$\lim_{k \to \infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = 0 \quad (15)$$

holds. By Lemma 3.1, we get (12).

2) For Goldstein rule. Let the step size α_k be the solution of (6) and (11).

By (11) and the mean value theorem, we have

$$-\alpha_k g(x_k + \alpha_k \tau_k d_k)^T d_k = f_k - f_{k+1} \leq -\sigma_2 \alpha_k g_k^T d_k ,$$

where $\tau_k \in (0,1)$, thus

$$g(x_k + \alpha_k \tau_k d_k)^T d_k \geq \sigma_2 g_k^T d_k .$$

Using Assumption (ii) again, we get

$$-(1 - \sigma_2) g_k^T d_k$$
$$\leq [g(x_k + \alpha_k \tau_k d_k) - g_k]^T d_k \leq \alpha_k L \|d_k\|^2 ,$$

which combining with (6), and use Assumption (i), we have (14) and (15), respectively. By Lemma 3.1, (12) holds.

3) For strong Wolf-Powell rule. Let the step size α_k be the solution of (6) and (7).

By (7), we have

$$\sigma_2 g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k \leq -\sigma_2 g_k^T d_k .$$

Similar to the proof of the above case. We can obtain (12) immediately.

4) For weak Wolf-Powell rule. Let the step size α_k be the solution of (6) and (8). Similar to the proof of the case 3), we can also get (12).

Then we conclude this result of this theorem.

By Lemma 3.1, there exists a constant $\delta_0 > 0$ such that

$$-\frac{g_k^T d_k}{\|g_k\| \|d_k\|} \geq \delta_0, \text{ for all } k. \quad (16)$$

By the proof process of Lemma 3.1. We can deduce that there exists a positive number δ_1 satisfying

$$f_k - f_{k+1} \geq \delta_1 \left(\frac{-g_k^T d_k}{\|d_k\|} \right)^2, \text{ for all } k. \quad (17)$$

Similar to the proof of Theorem 4.1 in [27], it is not difficult to prove the linear convergence rate of Algorithm 1. We state the theorem as follows but omit the proof.

Theorem 3.2 (*see* [27]) *Based on* (16), (17), *and the condition that the function* f *is twice continuously differentiable and uniformly convex on* R^n. *Let* $\{\alpha_k, d_k, x_{k+1}, g_{k+1}\}$ *be generated by Algorithm 1 with the exact minimization rule, the Goldstein line search rule, the SWP line search rule, or the WWP line search rule. Then* $\{x_k\}$ *converges to* x^* *at least linearly, where* x^* *is the unique minimal point of* $f(x)$.

4. Numerical Results

In this section, we report some numerical experiments with Algorithm 1, Algorithm 2, and Algorithm 3. We test these algorithms on some problem [28] taken from MATLAB with given initial points. The parameters common to these methods were set identically, $\sigma_1 = 0.1$, $\sigma_2 = 0.9$, $\varepsilon = 10^{-6}$, In this experiment, the following *Himmeblau* stop rule is used:

If $|f(x_k)| > e_1$, let $stop1 = \frac{|f(x_k) - f(x_{k+1})|}{|f(x_k)|}$; Otherwise, let $stop1 = |f(x_k) - f(x_{k+1})|$, where $e_1 = 10^{-5}$. If $\|g_k\| < \varepsilon$ or $stop1 < e_2$ was satisfied, the program will be stopped, where $e_2 = 10^{-5}$.

We also stop the program if the iteration number is more than one thousand. Since the line search cannot always ensure the descent condition $d_k^T g_k < 0$, uphill search direction may occur in the numerical experiments. In this case, the line search rule maybe failed. In order to avoid this case, the stepsize α_k will be accepted if the

searching number is more than forty in the line search. The detailed numerical results are listed on the web site
http://210.36.18.9:8018/publication.asp?id=34402

Dolan and Moré [29] gave a new tool to analyze the efficiency of Algorithms. They introduced the notion of a performance profile as a means to evaluate and compare the performance of the set of solvers S on a test set P. Assuming that there exist n_s solvers and n_p problems, for each problem p and solver s, they defined

$t_{p,s}$ = computing time (the number of function evaluations or others) required to solve problem p by solver s.

Requiring a baseline for comparisons, they compared the performance on problem p by solver s with the best performance by any solver on this problem; that is, using the performance ratio

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : s \in S\}}.$$

Suppose that a parameter $r_M \geq r_{p,s}$ for all p, s is chosen, and $r_{p,s} = r_M$ if and only if solver s does not solve problem p.

The performance of solver s on any given problem might be of interest, but we would like to obtain an overall assessment of the performance of the solver, then they defined

$$\rho_s(t) = \frac{1}{n_p} size\{p \in P : r_{p,s} \leq t\},$$

Thus $\rho_s(t)$ was the probability for solver $s \in S$ that a performance ratio $r_{p,s}$ was within a factor $t \in R$ of the best possible ration. Then function ρ_s was the (cumulative) distribution function for the performance ratio. The performance profile $\rho_s : R \in [0,1]$ for a solver was a nondecreasing, piecewise constant function, continuous from the right at each breakpoint. The value of $\rho_s(1)$ was the probability that the solver would win over the rest of the solvers.

According to the above rules, we know that one solver whose performance profile plot is on top right will win over the rest of the solvers.

In **Figures 1-3**, NA denotes Algorithm 1, PRP denotes Algorithm 2, and PRP$^+$ denotes Algorithm 3. **Figures 1-3** show that the performance of these methods is relative to $NT = NF + m * NG$, where NF and NG denote the number of function evaluations and gradient evaluations respectively, and m is an integer. According to the results on automatic differentiation [30], the value of m can be set to $m = 5$. That is to say, one gradient evaluation is equivalent to m number of function evaluations if automatic differentiation is used. From these three figures
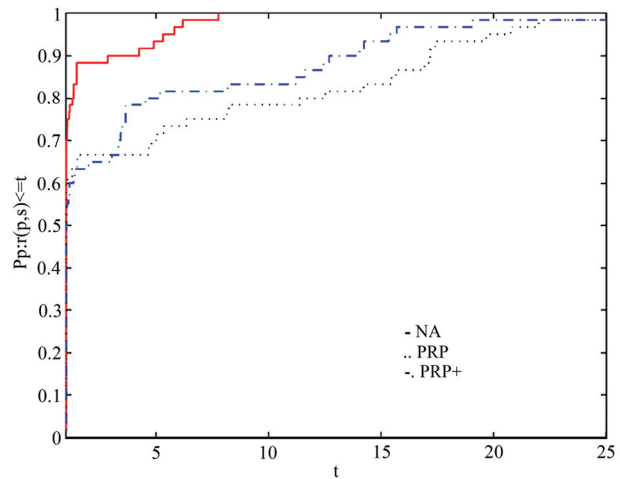

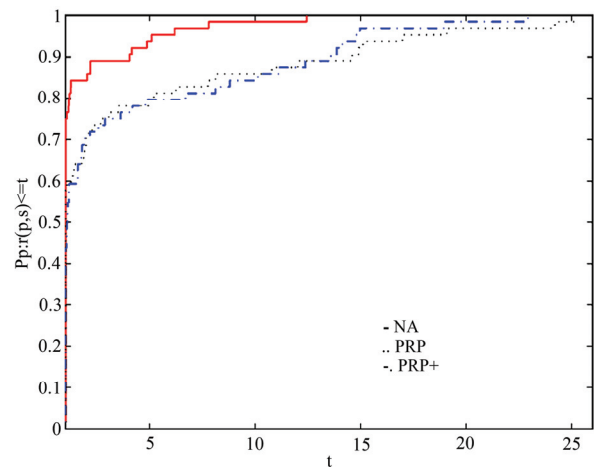**Figure 1. Performance profiles(NT) of methods with Goldstein rule**

Figure 2. Performance profiles(NT) of methods with strong Wolfe-Powell rule
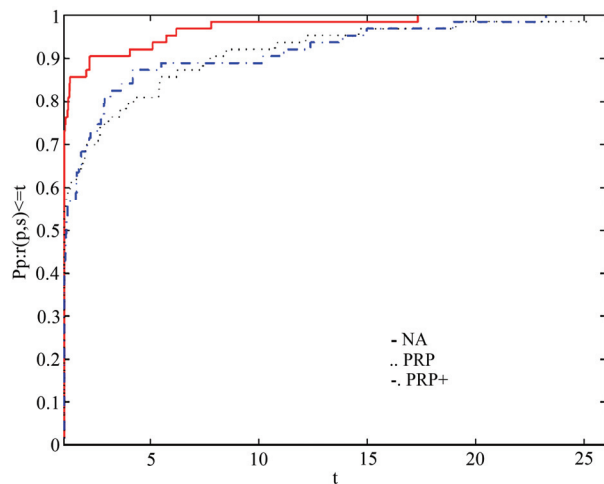

Figure 3. Performance profiles (NT) of methods with weak Wolfe-Power rule

it is clear that the given method has the most wins (has the highest probability of being the optimal solver).

In summary, the presented numerical results reveal that the new method, compared with the normal PRP method and the modified PRP method [14], has potential advantages.

5. Conclusions

This paper gives a new line search method for unconstrained optimization. The global and R-linear convergence are established under weaker assumptions on the search direction d_k. Especially, the direction d_k satisfies the sufficient condition (5) and the condition (9) without carrying out any line search technique, and some paper [14,27,30] often obtains these two conditions by assumption. The comparison of the numerical results shows that the new search direction of the new algorithm is a good search direction at every iteration.

REFERENCES

[1]   G. Yuan and X. Lu, "A New Line Search Method with Trust Region for Unconstrained Optimization," *Communications on Applied Nonlinear Analysis*, Vol. 15, No. 1, 2008, pp. 35-49.

[2]   G. Yuan, X. Lu, and Z. Wei, "New Two-Point Stepsize Gradient Methods for Solving Unconstrained Optimization Problems," *Natural Science Journal of Xiangtan University*, Vol. 29, No. 1, 2007, pp. 13-15.

[3]   G. Yuan and Z. Wei, "New Line Search Methods for Uncons- trained Optimization," *Journal of the Korean Statistical Society*, Vol. 38, No. 1, 2009, pp. 29-39.

[4]   Y. Yuan and W. Sun, "Theory and Methods of Optimization," Science Press of China, Beijing, 1999.

[5]   D. C. Luenerger, "Linear and Nonlinear Programming," 2nd Edition, Addition Wesley, Reading, MA, 1989.

[6]   J. Nocedal and S. J. Wright, "Numerical Optimization," Springer, Berlin, Heidelberg, New York, 1999.

[7]   Z. Wei, G. Li, and L. Qi, "New Quasi-Newton Methods for Unconstrained Optimization Problems," *Applied Mathematics and Computation*, Vol. 175, No. 1, 2006, pp. 1156-1188.

[8]   Z. Wei, G. Yu, G. Yuan, and Z. Lian, "The Superlinear Convergence of a Modified BFGS-type Method for Unconstrained Optimization," *Computational Optimization and Applications*, Vol. 29, No. 3, 2004, pp. 315-332.

[9]   G. Yuan and Z. Wei, "The Superlinear Convergence Analysis of a Nonmonotone BFGS Algorithm on Convex Objective Functions," *Acta Mathematica Sinica, English Series*, Vol. 24, No. 1, 2008, pp. 35-42.

[10] G. Yuan and Z. Wei, "Convergence Analysis of a Modified BFGS Method on Convex Minimizations," *Computational Optimization and Applications*, Science Citation Index, 2008.

[11] Y. Dai and Y. Yuan, "A Nonlinear Conjugate Gradient with a Strong Global Convergence Properties," *SIAM Journal of Optimization*, Vol. 10, No. 1, 2000, pp. 177-182.

[12] Z. Wei, G. Li, and L. Qi, "New Nonlinear Conjugate Gradient Formulas for Large-Scale Unconstrained Optimization Problems," *Applied Mathematics and Computation*, Vol. 179, No. 2, 2006, pp. 407-430.

[13] G. Yuan and X. Lu, "A Modified PRP Conjugate Gradient Method," *Annals of Operations Research*, Vol. 166, No. 1, 2009, pp. 73-90.

[14] J. C. Gibert, J. Nocedal, "Global Convergence Properties of Conjugate Gradient Methods for Optimization," *SIAM Journal on Optimization*, Vol. 2, No. 1, 1992, pp. 21-42.

[15] Y. Dai and Y. Yuan, "Nonlinear Conjugate Gradient Methods," Shanghai Science and Technology Press, 2000.

[16] E. Polak and G. Ribiè, "Note Sur la Xonvergence de Directions Conjugèes," *Rev Francaise Informat Recherche Operatinelle 3e Annèe*, Vol. 16, 1969, pp. 35-43.

[17] M. J. D. Powell, "Nonconvex Minimization Calculations and the Conjugate Gradient Method," *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, Vol. 1066, 1984, pp. 122-141.

[18] L. Grippo and S. Lucidi, "A Globally Convergent Version of the Polak-RibiÈRe Gradient Method," *Mathematical Programming*, Vol. 78, No. 3, 1997, pp. 375-391.

[19] W. W. Hager and H. Zhang, "A New Conjugate Gradient Method with Guaranteed Descent and an Efficient Line Search," *SIAM Journal on Optimization*, Vol. 16, No. 1, 2005, pp. 170-192.

[20] Z. Wei, S. Yao, and L. Liu, "The Convergence Properties of Some New Conjugate Gradient Methods," *Applied Mathematics and Computation*, Vol. 183, No. 2, 2006, pp. 1341-1350.

[21] G. H. Yu, "Nonlinear Self-Scaling Conjugate Gradient Methods for Large-scale Optimization Problems," Thesis of Doctor's Degree, Sun Yat-Sen University, 2007.

[22] G. Yuan, "Modified Nonlinear Conjugate Gradient Methods with Sufficient Descent Property for Large-Scale Optimization Problems," *Optimization Letters*, Vol. 3, No. 1, 2009, pp. 11-21.

[23] G. Yuan, "A Conjugate Gradient Method for Unconstrained Optimization Problems," *International Journal of Mathematics and Mathematical Sciences*, Vol. 2009, 2009, pp. 1-14.

[24] G. Yuan, X. Lu, and Z. Wei, "A Conjugate Gradient Method with Descent Direction for Unconstrained Optimization," *Journal of Computational and Applied Mathematics*, Vol. 233, No. 2, 2009, pp. 519-530.

[25] L. Zhang, W. Zhou, and D. Li, "A Descent Modified Polak-RibiÈRe-Polyak Conjugate Method and its Global Convergence," *IMA Journal on Numerical Analysis*, Vol. 26, No. 4, 2006, pp. 629-649.

[26] Y. Liu and C. Storey, "Efficient Generalized Conjugate Gradient Algorithms, Part 1: Theory," *Journal of Optimization Theory and Application*, Vol. 69, No. 1, 1992, pp. 17-41.

[27] Z. J. Shi, "Convergence of Line Search Methods for Unconstrained Optimization," *Applied Mathematics and Computation*, Vol. 157, No. 2, 2004, pp. 393-405.

[28] J. J. Moré, B. S. Garbow, and K. E. Hillstrome, "Testing Unconstrained Optimization Software," *ACM Transactions on Mathematical Software*, Vol. 7, No. 1, 1981, pp. 17-41.

[29] E. D. Dolan and J. J. Moré, "Benchmarking Optimization Software with Performance Profiles," *Mathematical Programming*, Vol. 91, No. 2, 2002, pp. 201-213.

[30] Y. Dai and Q. Ni, "Testing Different Conjugate Gradient Methods for Large-scale Unconstrained Optimization," *Journal of Computational Mathematics*, Vol. 21, No. 3, 2003, pp. 311-320.