ISSN 2277-3061

Two Modified Hager and Zhang's Conjugate Gradient Algorithms For Solving Large-Scale Optimization Problems

Yong Li, Gonglin Yuan
Department of Mathematics, Baise University, Baise, Guangxi, 533000, P. R. China.
yli909@126.com
Corresponding Author, College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, P.R. China
glyuan@gxu.edu.cn

ABSTRACT

At present, the conjugate gradient (CG) method of Hager and Zhang (Hager and Zhang, SIAM Journal on Optimization, 16(2005)) is regarded as one of the most effective CG methods for optimization problems. In order to further study the CG method, we develop the Hager and Zhang's CG method and present two modified CG formulas, where the given formulas possess the value information of not only the gradient but also the function. Moreover, the sufficient descent condition will be holden without any line search. The global convergence is established for nonconvex function under suitable conditions. Numerical results show that the proposed methods are competitive to the normal conjugate gradient method.

Indexing terms/Keywords

conjugate gradient; sufficient descent; global convergence.

Academic Discipline And Sub-Disciplines

Operation research;Nonlinear programming.

SUBJECT CLASSIFICATION

90C26.

TYPE (METHOD/APPROACH)

Algorithm design and numerical experiment.

Council for Innovative Research

Peer Review Research Publishing System

Journal: International Journal of Computers & Technology

Vol 11, No.5

editor@cirworld.com

www.cirworld.com, member.cirworld.com

1 INTRODUCTION

Consider the following unconstrained optimization problem

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

where $f:\Re^n \to \Re$ is continuously differentiable. The nonlinear conjugate gradient method is one of the most effective line search methods for (1.1) because of its simplicity and its very low memory requirement. This method can avoid, like steepest descent method, the computation and storage of some matrices associated with the Hessian of objective functions. The following iterative formula is often used by CG method

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

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

$$d_{k+1} = \begin{cases} -g_{k+1} + \beta_k d_k, if \ k \ge 1 \\ -g_{k+1}, \quad if \ k = 0, \end{cases} \tag{1.3}$$

where g_k is the gradient of $f(x)$ at the point x_k , and $\beta_k \in \Re$ is a scalar which determines the different conjugate gradient methods. These based conjugate gradient methods [12, 17, 18, 27, 35] are equivalent (see [13, 48] etc) in the linear case, namely, when f is a strictly convex quadratic function and α_k is calculated by the following exact minimization rule: the step size α_k is chosen such that

$$f(x_k + \alpha_k d_k) = \min_{\alpha\ge 0} f(x_k + \alpha d_k) \tag{1.4}$$

One of the most efficient formula for β_k is the following PRP method [35]

$$\beta_k^{PRP} = \frac{g_{k+1}^T(g_{k+1} - g_k)}{\|g_k\|^2}, \tag{1.5}$$

where g_k and g_{k+1} are the gradients $\nabla f(x_k)$ and $\nabla f(x_{k+1})$ of $f(x)$ at the point x_k and x_{k+1} , respectively, and $\|.\|$ denotes the Euclidian norm of vectors. For its convergent results, Polak and Ribiere [35] presenteded the global convergence with the exact line search for convex functions. Powell [36] gave a counter example to show that there exist nonconvex functions on which the PRP method does not converge globally even using the exact line search. He suggested that β_k should not be less than zero, which is very important to ensure the global convergence (see [13, 37]). Considering the above suggestion, Gilbert and Nocedal[19] proved that the modified PRP method $\beta_k^+ = \max\ 0, \beta_k^{PRP}$ is globally convergent with the WWP line search under the assumption of sufficient descent condition.

From the literature, one hopes to find the steplength α_k using the following weak Wolfe-Powell (WWP) line search

$$f(x_k + \alpha_k d_k) \le f_k + \delta\alpha_k g_k^T d_k \tag{1.6}$$

And

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{1.7}$$

where $\delta \in (0,1/2)$, and $\sigma \in (\delta,1)$. However, the global convergence of the PRP conjugate gradient method is still open with the above WWP conditions. Some formulas which possess the global convergence property (such as β_k^{DY} [12]) with the WWP did not perform better than the performances of the PRP method in numerical computation. Based β_k^{DY} , Dai and Yuan [14] use the WWP condition and propose an efficient conjugate gradient method. 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 [13, 19]). Thus, any new conjugate gradient method should at least satisfy one of the following conditions [41]:

(i) The method with the WWP line search rule (or other line search rules) has some strongly convergent properties, at least, the method with the formula and the WWP line search rule (or other line search rules) may generate a descent direction at each iteration, and converges globally.

(ii) The average performances on the numerical computation of the formula with WWP line search rule (or others)should not be much inferior to the ones of the PRP.

The following sufficient descent condition

$$g_k^T d_k \le -c\|g_k\|^2, \ \forall\, k \ge 0 \ and \ some \ cons\tan t \ c > 0 \tag{1.8}$$

is often used to analyze the global convergence of the nonlinear conjugate gradient method with the inexact line search techniques. Toouati-Ahmed and Storey [1], Al-Baali [2], Gilbert and Nocedal [19],and Hu and Storey [28] hinted that the sufficient descent condition may be crucial for conjugate gradient methods. In order to ensure the sufficient descent condition and establish the convergence of the PRP method, Grippo and Lucidi [21] presented a new line search rule. Resent years, some good results on the nonlinear conjugate gradient method are given (see [3, 9, 25, 42]). But for some methods which have been studied in the optimization area, such as the steepest descent method and the Newton method,

the descent properties or the sufficient descent properties are independent of line searches. Is there any nonlinear conjugate gradient formula which possesses the sufficient descent property (1.8) without any line search? Many authors answer this question positively (see [21, 23, 24, 43, 44, 45, 46, 47, 49, 53, 56]etc.). For instance, Zhang, Zhou, and Li [56] presented a modified PRP method with

$$d_{k+1} = \begin{cases} -g_{k+1} + \beta_k^{PRP} d_k - \vartheta_k y_k & if \ k \ge 1 \\ -g_{k+1}, & if \ k = 0, \end{cases} \tag{1.9}$$

where $\vartheta_k = \dfrac{g_{k+1}^T d_k}{\|g_k\|^2}$. It is not difficult to get $d_k^{\ T} g_k = -\|g_k\|^2$. This method can reduces to a standard PRP method if exact line search is used, its global convergence with Armijo-type line search is obtained, but fails to WWP line search. Based on [11], Hager and Zhang proposed a new conjugate gradient method [23]

$$d_{k+1} = \begin{cases} -g_{k+1} + \beta_k^{HZ} d_k & if \ k \ge 1 \\ -g_{k+1}, & if \ k = 0, \end{cases} \tag{1.10}$$

where $\beta_k^{HZ} = \dfrac{g_{k+1}^T\left(y_k - 2\dfrac{\|y_k\|^2}{s_k^T y_k} s_k\right)}{d_k^T y_k}$, $s_k = x_{k+1} - x_k, and \ y_k = g_{k+1} - g_k$. This method can guarantee that d_k provides a descent direction of f at x_k . Moreover, d_k satisfies $d_k^T g_k \le -\dfrac{7}{8}\|g_k\|^2$. This method can be regarded as a modified HS method and possess global convergence with WWP line search. Furthermore,they gave another more effective CG formula [24]

$$\beta_k^{HZ^+} = \max \ \beta_k^{HZ}, \varsigma_k \ ,$$

where $\varsigma_k = \dfrac{-1}{\|d_k\|, \min \ \varsigma, \|g_k\|}$ and $\varsigma > 0$ is a constant. Numerical results show that this method is better than the others conjugate gradient methods (such as the PRP, the PRP+, the HS, and the DY, etc.)and the limited memory BFGS method (see [23, 24] in detail). Today, this method (1.10) is considered to be one of the most effective algorithm. Therefore, any new conjugate gradient method should satisfy the following two conditions:

(j) The method with the WWP line search rule (or other line search rules) has some strongly convergent properties, at least, the method with the formula and the WWP line search rule (or other line search rules) may generate a sufficient descent direction at each iteration, and converges globally.

(jj) The average performances on the numerical computation of the formula with WWP line search rule (or others) should not be much inferior to the ones of the HZ^+ .

In this paper, we design two new CG formulas to satisfy the above conditions. The main attributes of this paper are as follows.

- The given method possesses the the sufficient descent property without any line search technique.
- The global convergence of the new methods is established for nonconvex function.
- Numerical results show that these two methods are competitive to the HZ^{+} method.

In the next section, motivation and algorithm are stated. The sufficient descent property and the global convergence of the new method are proved in Section 3. In the Section 4, the numerical results are reported. One conclusion is stated in the last section.

2. MOTIVATION AND ALGORITHM

In this section, we will give motivations based on the BFGS formulas and the line search technique,respectively.

2.1 Motivations based on BFGS formula

It is well known that the BFGS method is one of the most effective methods for unconstrained optimization problems. There are many good results can be found (see [5, 6, 7, 8, 15, 29, 30, 32, 49] etc.). Where Wei,Yu, Yuan, and Lian [40] presented a new BFGS update method generated by Taylor's formula as follows:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k^* y_k^{*T}}{s_k^T y_k^*}, \tag{2.1}$$

where $y_k^* = y_k + \frac{\rho_k}{\|s_k\|^2} s_k$ and $\rho_k = 2\ f(x_k) - f(x_k + \alpha_k d_k)\ + (g(x_k + \alpha_k d_k) + g(x_k))^T s_k$ . . Under the assumption that the objective functions are uniformly convex ones, the superlinear convergence of the new BFGS algorithm was given with the weak Wolfe-Powell (WWP) linesearch. Observing the quasi-Newton Equation

$$B_{k+1} s_k = y_k^* \tag{2.2}$$

which contains not only gradient value information but also function value information at the present and the previous step, one may argue that the resulting methods will really outperform than the original method. In fact, the practical computation shows that this method is better than the normal BFGS method (see [39, 40] for detail). Furthermore, some theoretical advantages of the new quasi-Newton equation (2.2) can be seen from the following two theorems.

Theorem 2.1 (Lemma 3.1 [40]) Considering the quasi-Newton equation (2.2). Then we have for all k≥1

$$f(x_k) = f(x_{k+1}) + g(x_{k+1})^T (x_k - x_{k+1}) + \frac{1}{2}(x_k - x_{k+1})^T B_{k+1}(x_k - x_{k+1}).$$

Theorem 2.2 (Theorem 3.1 [31]) Assume that the function $f(x)$ is suffciently smooth and $\|s_k\|$ is suffciently small, then we have

$$s_k^T G_{k+1} s_k - s_k^T y_k^* - \frac{1}{3} s_k^T (T_{k+1} s_k) s_k = o(\|s_k\|^4) \tag{2.3}$$

and

$$s_k^T G_{k+1} s_k - s_k^T y_k - \frac{1}{2} s_k^T (T_{k+1} s_k) s_k = o(\|s_k\|^4) \tag{2.4}$$

where G_{k+1} denotes the Hessian matrix of f at x_{k+1} , T_{k+1} is the tensor of f at x_{k+1} , and

$$s_k^T (T_{k+1} s_k) s_k = \sum_{i,j,l=1}^{n} \frac{\partial^3 f(x_{k+1})}{\partial x^i \partial x^j \partial x^l} s_k^i s_k^j s_k^l .$$

It can be seen that if the objective function f is uniformly convex, then

$$s_k^T y_k^* = s_k^T y_k + 2\ f_k - f_{k+1}\ +\ g_{k+1} + g_k\ ^T s_k = 2 s_k^T g_{k+1} + 2(f_k - f_{k+1}) > 0$$

holds, where the last inequality is from the uniform convexity of f . Hence, the update formula (2.1)can ensure the positive definiteness of the matrix B_k for uniformly convex function, and the superlinear convergence of this method has been established. However, if f is a general convex function, then $s_k^T y_k^*$ may equal to 0. In this case, the positive definiteness of the update matrix B_k can not be sure.Moreover, the global convergence and the superlinear convergence are still open for the general convexfunction. Motivated by the above observations, we study whether there exists another quasi-Newtonformula whose approximation for the Hessian of the objective function is not inferior to those of the formula (2.1) or the normal BFGS formula in some sense, which possesses the global convergence and the superlinear convergence for general convex function and its numerical results are competitive to those of other similar methods. Now we discuss ρ_k for general convex functions in the following two cases:

case i: If $\rho_k > 0$ we have

$$s_k^T (y_k + \frac{\rho_k}{\|s_k\|^2} s_k) = s_k^T y_k + \rho_k > s_k^T y_k \geq 0 \ . \tag{2.5}$$

case ii: If $\rho_k < 0$, we get

$$\begin{aligned} 0 > \rho_k &= 2\ [f(x_k) - f(x_k + \alpha_k d_k)]\ + (g(x_k + \alpha_k d_k) + g(x_k))^T s_k \\ &\geq -2 g_{k+1}^T s_k + (g(x_k + \alpha_k d_k) + g(x_k))^T s_k \\ &= -s_k^T y_k , \end{aligned} \tag{2.6}$$

which means that $s_k^T y_k > 0$ holds. Therefore, we define the quasi-Newton equation as follows [or see [50]in detail]:

$$B_{k+1} s_k = y_k^m \ , \tag{2.7}$$

where $y_k^m = y_k + \frac{\max\{\rho_k, 0\}}{\|s_k\|^2} s_k$,and

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k^m (y_k^m)^T}{s_k^T y_k^m} , \tag{2.8}$$

which can ensure B_{k+1} inherits the positive definiteness of B_k for the general convex function. The global convergence and the superlinear convergence have been established for general convex functions.Numerical results show that this method is interesting.Zhang, Deng, and Chen [54] presented the following quasi-Newton equation:

$$B_{k+1} s_k = y_k^{3*} = y_k + \overline{A_k} s_k \ , \tag{2.9}$$

Where

$$\overline{A_k} = \frac{6\ [f(x_k) - f(x_{k+1})]\ + 3\ [g(x_{k+1}) + g(x_k)]^T s_k}{\|s_k\|^2} \ . \tag{2.10}$$

They obtained the following modified BFGS-type update formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k^{3*} y_k^{3*T}}{y_k^{3*T} s_k} \ .$$

This quasi-Newton equation (2.9) contains both gradient and function value information at the current and the previous step too, one may argue that the resulting methods will really outperform than the original method. In fact, the practical computation shows that this method is better than the normal BFGS method (see [54] for detail). Furthermore, some theoretical advantages of the new quasi-Newton equation (2.9) can be seen from the following theorem.

Theorem 2.3 (Theorem 3.3 [54]) Assume that the function $f(x)$ is suffciently smooth and $\|s_k\|$ is suffciently small, then we have

$$s_k^T(G_{k+1}s_k - y_k^{3*}) = o(\|s_k\|^4) \tag{2.11}$$

and

$$s_k^T(G_{k+1}s_k - y_k) = o(\|s_k\|^3)\,. \tag{2.12}$$

Similarly, we can define another quasi-Newton equation as follows

$$B_{k+1}s_k = y_k^{mm}\,,$$

where $y_k^{mm} = y_k + \max\ \overline{A_k}, 0\ s_k$. It is not diffcult to deduce that $s_k^T y_k^{mm} > 0$ holds for generally convexk functions.

Motivated by the above discussions and the conjugate gradient method HZ, the modified HZ formulas are to replace y_k by y_k^m and y_k^{mm} respectively. Namely the new conjugate gradient formulas are defined by

$$d_{k+1} = \begin{cases} -g_{k+1} + \beta_k^m d_k\,, & if\ k \ge 1 \\ -g_{k+1} & if\ k = 0, \end{cases} \tag{2.13}$$

and

$$d_{k+1} = \begin{cases} -g_{k+1} + \beta_k^{mm} d_k\,, & if\ k \ge 1 \\ -g_{k+1} & if\ k = 0, \end{cases} \tag{2.14}$$

where $\beta_k^m = \dfrac{g_{k+1}^T\left(y_k^m - 2\dfrac{\|y_k^m\|^2}{s_k^T y_k^m} s_k\right)}{d_k^T y_k^m}$ .and $\beta_k^{mm} = \dfrac{g_{k+1}^T\left(y_k^{mm} - 2\dfrac{\|y_k^{mm}\|^2}{s_k^T y_k^{mm}} s_k\right)}{d_k^T y_k^{mm}}$ .In the following section,we state our algorithms.

2.2 Algorithms

The earliest nonmonotone line search framework was developed by Grippo, Lampariello, and Lucidi in[20] for Newton's methods. Many subsequent papers have exploited nonmonotone line search techniques of this nature (see [4, 26, 33, 51, 52, 57] etc.). Although these nonmonotone technique work well in many cases, there are some drawbacks. First, a good function value generated in any iteration is essentially discarded due to the max in the nonmonotone line search technique. Second, in some cases,the numerical performance is very dependent on the choice of M (see [20, 38]). Zhang and Hager [55]presented a new nonmonotone line search technique defined by:

$$f(x_k + \alpha_k d_k) \le C_k + \delta\alpha_k g(x_k)^T d_k\,, \tag{2.15}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g(x_k)^T d_k \quad , \tag{2.16}$$

where $0 < \delta < \sigma < 1$, $C_{k+1} = \dfrac{\eta_k Q_k C_k + f(x_k + \alpha_k d_k)}{Q_{k+1}}$, $Q_{k+1} = \eta_k Q_k + 1$, $\eta_k \in \eta_{\min}, \eta_{\max}$, $0 \le \eta_{\min} \le \eta_{\max} \le 1$, $C_1 = f(x_1)$,and $Q_1 = 1$. It is not difficult to see that C_{k+1} is a convex combination of C_k and $f(x_{k+1})$. Since $C_1 = f(x_1)$, it follows that C_k is a convex combination of the function values $f(x_1), f(x_2), \cdots, f(x_k)$. The choice of η_k controls the degree of nonmonotonicity. If $\eta_k = 0$ for each k , then the line search is the usual monotone Wolfe or Armijo line search. If $\eta_k = 1$ for each k, then $C_k = A_k$, where

$$A_k = \frac{1}{k}\sum_{i=1}^{k} f(x_i)$$

is the average function value. Numerical results show that this technique is better than the normal nonmonotone technique. Considering the efficiency of this technique, we will use this technique to find steplength α_k . Based on the above discussions, we state our algorithms as follows.

Algorithm 1(New-cg1)

Step 0: Choose an initial point $x_1 \in \Re^n, \varepsilon \in (0,1), 0<\delta<\sigma<1, 0 \le \eta_{\min} \le \eta_{\max} \le 1.$ Set $d_1 = -g_1 = -\nabla f(x_1), Q_1 = 1, C_1 = f(x_1), k := 1.$

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

Step 2: Compute step size α_k by line search rules (2.15) and (2.16).

Step 3: let $x_{k+1} = x_k + \alpha_k d_k$. If $\|g_{k+1}\| \le \varepsilon$, then stop.

Step 4: Calculate the search direction by (2.13)

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

Algorithm 2(New-cg2)

Step 4 of Algorithm 1 is replaced by: Calculate the search direction by (2.14).

In the following section, we will show that the given two algorithms possess the sufficiently descent property without any line search technique and the global convergence for the general functions.

3. THE SUFFICIENT DESCENT PROPERTY AND THE GLOBAL CONVERGENCE

With conjugate gradient methods, the line search typically requires sufficient accuracy to ensure that the search directions yield descent [10, 22], Moreover, it has been shown [12] that for the Fletcher-Reeves [18] and Polak-Ribie`-Polyak [34, 35] conjugate gradient methods, a line search that satisfies the strong Wolfee conditions may not yield a direction of descent for a suitable choice the Wolfe line search parameters, even for the function $f(x) = \lambda\|x\|^2$, where $\lambda > 0$ is a constant. An attractive feature of these two conjugate gradient is that the search directions always yield descent.

Lemma 3.1 Consider (2.13) and (2.14), we have

$$g_{k+1}^T d_{k+1} \le -\frac{7}{8}\|g_{k+1}\|^2 . \tag{3.1}$$

Proof. Since $d_1 = -g_1$, we get $g_1^T d_1 = -\|g_1\|^2$, then (3.1) holds. For $k \ge 1$, multiplying (2.13) by g_{k+1}^T ,we obtain

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_k^m g_{k+1}^T d_k \\ &= -\|g_{k+1}\|^2 + g_{k+1}^T d_k \left(\frac{g_{k+1}^T y_k^m}{d_k^T y_k^m} - 2\frac{\|y_k^m\|^2 g_{k+1}^T d_k}{(d_k^T y_k^m)^2}\right) \\ &= \frac{g_{k+1}^T y_k^m d_k^T y_k^m g_{k+1}^T d_k - \|g_{k+1}\|^2 (d_k^T y_k^m)^2 - 2\|y_k^m\|^2 (g_{k+1}^T d_k)^2}{(d_k^T y_k^m)^2} \end{aligned} \tag{3.2}$$

Let $u = \frac{1}{2}(d_k^T y_k^m) g_{k+1}, v = 2(g_{k+1}^T d_k) y_k^m$, and use the inequality $v^T u \le \frac{1}{2}(\|v\|^2 + \|u\|^2)$,we have

$$g_{k+1}^T y_k^m d_k^T y_k^m g_{k+1}^T d_k \le \frac{1}{2}\left(\left\|\frac{1}{2} d_k^T y_k^m g_{k+1}\right\|^2 + \left\|2 g_{k+1}^T d_k y_k^m\right\|^2\right)$$

$$= \frac{1}{8}(d_k^T y_k^m)^2 \|g_{k+1}\|^2 + 2(g_{k+1}^T d_k)^2 \|y_k^m\|^2 , \tag{3.3}$$

this implies that $g_{k+1}^T d_{k+1} \le -\frac{7}{8}\|g_{k+1}\|^2$ by considering the last equality of (3.2). Similarly, we can also get (3.1) from (2.14). The proof is complete.

In the following we assume that $g_k \ne 0$ for all k, for otherwise a stationary point has been found.The following assumptions are often used to prove the convergence of the nonlinear conjugate gradient methods (see [18, 27, 35, 43, 49] etc.).

Assumption 3.1 (i) The level set $\Omega = \ x \in \Re^n \,|\, f(x) \le 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 constant $L > 0$ such that

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

Lemma 3.2 Suppose that Assumption 3.1 holds. Let the sequence g_k and d_k be generated by Algorithm 1. Then

$$\alpha_k \ge \frac{1-\sigma}{L}\frac{|g_k^T d_k|}{\|d_k\|^2}, \tag{3.5}$$

and

$$s_k^T y_k^m \ge \frac{7(1-\sigma)}{8}\|g_k\|^2 , \tag{3.6}$$

$$\|y_k^m\| \le 2L\|s_k\| \tag{3.7}$$

hold.

Proof. By (2.16) and the Lipschitz condition (3.4), we have

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

Considering (3.1), we get (3.5). Using the definition of y_k^m , (3.1), and the relation (2.16), we obtain

$$d_k^T y_k^m = d_k^T\left(y_k + \frac{\max\ \rho_k, 0\ }{\|s_k\|^2} s_k\right) \ge d_k^T y_k = d_k^T (g_{k+1} - g_k) \ge -(1-\sigma) g_k^T d_k$$

$$> \frac{7(1-\sigma)}{8}\|g_k\|^2 \tag{3.8}$$

then (3.6) holds. In the following, we prove that (3.7) holds. By mean value theorem, we get

$$\begin{aligned}
\rho_k &= 2(f_k - f_{k+1}) + (g_{k+1} + g_k)^T s_k \\
&= (-2g(x_k + \theta s_k) + g_{k+1} + g_k)^T s_k \\
&\le \|s_k\| \left[\|g_{k+1} - g(x_k + \theta s_k)\| + \|g_k - g(x_k + \theta s_k)\| \right] \\
&\le \|s_k\| \left[L(1-\theta)\|s_k\| + L\theta \|s_k\| \right] \\
&= L\|s_k\|^2 ,
\end{aligned} \tag{3.9}$$

where $\theta \in (0,1)$ and the last inequality follows (3.4). Therefore, from the definition of y_k^m and the Lipschitz condition (3.4), we have

$$\|y_k^m\| = \left\| y_k + \frac{\max\{\rho_k, 0\}}{\|s_k\|^2} s_k \right\| \le \|y_k\| + \frac{|\rho_k|\|s_k\|}{\|s_k\|^2} \le 2L\|s_k\|.$$

Then (3.7) holds. This completes the proof。

Based on Assumption 3.1, Lemma 3.1, and Lemma 3.2, similar to the Theorem 3.2 in [23], it is not difficult to prove the following global convergence theorem of Algorithm 1. So we only state it as follows,but omit the proof。

Theorem 3.1 Let Assumption 3.1 hold and the sequence $\{g_k, d_k\}$ be generated by Algorithm 1. Then

$$\liminf_{k\to\infty} \|g_k\| = 0.$$

In a way similar to the above discussions, we can also get the global convergence of Algorithm 2. In this paper, we do not prove it anymore.

4. NUMERICAL RESULTS

In this section, we test the numerical behavior of Algorithm 1. The algorithm is implemented by Fortran code in double precision arithmetic. All experiments are run on a PC with CPU Intel Pentium Dual E7500 2.93GHz, 2G bytes of SDRAM memory, and Red Hat Linux 9.03 operating system. Our experiments are performed on the subset of the nonlinear unconstrained problems from the CUTEr [17] collection, and the second-order derivatives of all the selected problems are available. Since we are interested in large problems,we refined this selection by considering only problems where the number of variables is at least 50. Altogether,we solved 72 problems. The names and characters of these problems are listed in Table 4.1.

TABLE 4.1(Test problems and their character)

Problems	Character
ARGLINA,ARGLINB,ARGLINC,BDQRTIC,BROWNAL,BROYDN7D,BRYBND CHAINWOO,CHNROSNB,COSINE,CRAGGLVY,CURLY10,CURLY20,DIXMAANA, DIXMAANB,DIXMAANC,DIXMAAND,DIXMAANE,DIXMAANF,DIXMAANG,DIXMAANH DIXMAANI,DIXMAANJ,DIXMAANL,DIXON3DQ,DQDRTIC,DQRTIC,EDENSCH EG2,ENGVAL1,ERRINROS,EXTROSNB,FLETCBV2,FLETCHCR,FREUROTH GENHUMPS,GENROSE,INDEF,LIARWHD,MANCINO,MSQRTALS,MSQRTBLS NONCVXU2,NONCVXUN,NONDIA,NONDQUAR,PENALTY1,PENALTY2,POWELLSG POWER,QUARTC,SCHMVETT,SENSORS,SINQUAD,SPARSINE,SPARSQUR SPMSRTLS,SROSENBR,TESTQUAD,TOINTGSS,TQUARTIC,TRIDIA VARDIM,VAREIGVL,WOODS	Academic
DECONVU,FMINSRF2,FMINSURF,MOREBV,TOINTGOR,TOINTQOR	Modelling

The program will be stopped when $\|g_k\|_\infty \le \max\{10^{-6}, 10^{-12}\|g_1\|_\infty\}$ was satisfied. The parameters and the line search rules are similar to [24]: $\delta = 0.1, \sigma = 0.9, \eta = 0.01$. The detailed numerical results are listed on Appendix I.

Dolan and More´ [16] gave a new tool to analyze the efficiency of Algorithms. They introduced the notion of a performance profile as 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

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

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

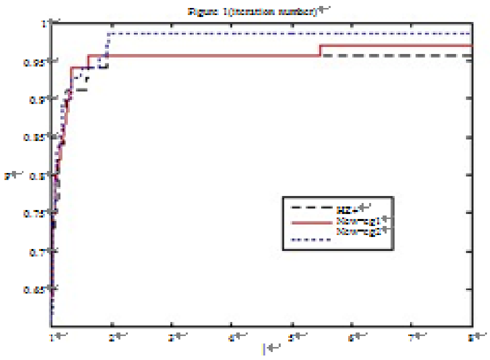
The performance of solver s on any given problem might be of interest, but we would like to obtain an overall assessment of the performance of the solver, then they defined

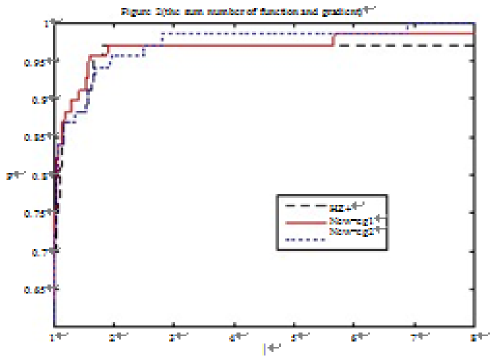
$$\rho_s(t) = \frac{1}{n_p} size \{ p \in P : \gamma_{p,s} \leq t \},$$

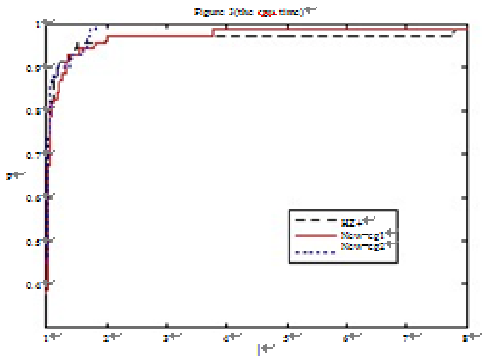
thus $\rho_s(t)$ was the probability for solver $s \in S$ that a performance ratio $\gamma_{p,s}$ was within a factor $t \in \Re$ of the best possible ration. Then function ρ_s was the (cumulative) distribution function for the performance ratio. The performance profile $\rho_s : \Re \mapsto [0,1]$ for a solver was a nondecreasing, piecewise constant function, and 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 rightwill win over the rest of the solvers.

In these three figures, HZ^+ denotes the algorithm in [24], New-cg1 denotes Algorithm 1, and New-cg2 denotes Algorithm 2, respectively. In Figure 1, 2, and 3, the performance denotes the iteration number,the number of function value and the gradient value, and the cpu time, respectively. From these three figures, it not difficult to see that Algorithm 2 perform best among these three algorithms, and Algorithm1 is competitive to the algorithm of HZ^+ .







5. CONCLUSION

In this paper, we propose two modified conjugate gradient formulas based on the well-known HZ formula,which possesses the sufficient descent condition without carrying out any line search too. The global convergence is established for nonconvex functions. Numerical results show that these two proposed methods are competitive to HZ^+ method. These two formulas have not only the gradient value information but also the function value information, moreover their quasi-Newton equation is closer to the Hessian matrix of the objective function than the normal quasi-Newton equation. This maybe make them possess better numerical results.

6. Acknowledgment. This work is supported by Guangxi Education researchproject (Grant No. 201010LX288), Guangxi NSF (Grant No. 2012GXNSFAA053002) and China NSF (Grant No. 11261006, 11161003 and 71001015).

REFERENCES

[1] T. Ahmed and D. Storey, Efficient hybrid conjugate gradient techniques, Journal of Optimization Theory and Applications, 64(1990), pp. 379-394.

[2] A. Al-Baali, Descent property and global convergence of the Flecher-Reeves method with inexact line search, IMA Journal of Numerical Analysis, 5(1985), pp. 121-124.

[3] N. Andrei, A hybrid conjugate gradient algorithm for unconstrained optimization as a convex combination of Hestenes-Stiefel and Dai-Yuan, Studies in Informatics and Control, 17(2008), pp. 55-70.

[4] E. G. Birgin, J. M. Martinez, and M. Raydan, Nonmonotone spectral projected gradient methods on convex sets, SIAM Journal on Optimization, 10(2000), pp. 1196-1121.

[5] C. G. Broyden, J. E. Dennis, and J. J. Mor´, On the local and supelinear convergence of quasi-Newton methods, Journal of the Institute of Mathematics and its Applications, 12(1973), pp. 223- 246.

[6] R. Byrd and J. Nocedal, A tool for the analysis of quasi-Newton methods with application to unconstrained minimization, SIAM Journal on Numerical Analysis, 26(1989), pp. 727-739.

[7] R. Byrd, J. Nocedal, and Y. Yuan, Global convergence of a class of quasi-Newton methods on convex problems, SIAM Journal on Numerical Analysis, 24(1987), pp. 1171-1189.

[8] Y. Dai, Convergence properties of the BFGS algorithm, SIAM Journal on Optimization, 13(2003), pp. 693-701.

[9] Y. Dai, A nonmonotone conjugate gradient algorithm for unconstrained optimization, Journal of Systems Science and Complexity, 15(2002), pp. 139-145.

[10] Y. Dai, J. Han, D. Sun, H. Yin, and Y. Yuan, Convergence properties of nonlinear conjugate gradient methods, SIAM Journal on Optimization, 10(1999), pp. 345-358.

[11] Y. Dai and L. Z. Liao, New conjugacy conditions and related nonlinear conjugate methods, Applied Mathematics and Optimization, 43(2001), pp. 87-101.

[12] Y. Dai and Y. Yuan, A nonlinear conjugate gradient with a strong global convergence properties, SIAM Journal on Optimization, 10 (2000), pp. 177-182.

[13] Y. Dai and Y. Yuan, Nonlinear conjugate gradient Methods, Shanghai Scientific and Technical Publishers, 1998.

[14] Y. Dai and Y. Yuan, An efficient hybrid conjugate gradient method for unconstrained optimization,Annals of Operations Research, 103(2001), pp. 33-47.

[15] J. E. Dennis and J. J. More´, A characteization of superlinear convergence and its application toequasi-Newton methods, Mathematics of Computation, 28(1974), pp. 549-560.

[16] E. D. Dolan and J. J. More´, Benchmarking optimization software with performance profiles, Math-ematical Programming, 91(2002), pp. 201-213.

[17] R. Fletcher, Practical method of optimization, Vol I: Unconstrained Optimization, 2nd edition, Wiley, New York, 1997.

[18] R. Fletcher and C. Reeves, Function minimization by conjugate gradients, The Computer Journal, 7(1964), pp. 149-154.

[19] J. C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM Journal on Optimization, 2(1992), pp. 21-42.

[20] L. Grippo, F. Lampariello, and S. Lucidi, A nonmonotone line search technique for newton's method,SIAM Journal on Numerical Analysis, 23(1986), pp. 707-716.

[21] L. Grippo and S. Lucidi, A globally convergent version of the Polak-Ribi`re gradient method, Math- ematical Programming, 78(1997), pp. 375-391.

[22] W. W. Hager, A derivate-based bracketing scheme for univariate minimization and the conjugate gradient method, Computers and Mathematics with Applications, 18(1989), pp. 779-795.

[23] W. W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM Journal on Optimization, 16(2005), pp. 170-192.

[24] W. W. Hager and H. Zhang, Algorithm 851: CGD ESCEN T, A conjugate gradient method with guaranteed descent, ACM Transactions on Mathematical Software, 32(2006), pp. 113-137.

[25] W. W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, Pacific Journal of Optimization, 2(2006), pp. 35-58.

[26] J. Y. Han and G. H. Liu, Global convergence Analysis of a New Nonmonotone BFGS Algorithm on Convex Objective Functions, Computational Optimization and Applications, 7(1997), pp. 277-289.

[27] M. R. Hestenes and E. Stiefel, Method of conjugate gradient for solving linear equations, Journal of Research of the National Bureau of Standards, 49(1952), pp. 409-436.

[28] Y. F. Hu and C. Storey, Global convergence result for conjugate method, Journal OptimizationTheory and Applications, 71(1991), pp. 399-405.

[29] D. Li and M. Fukushima, A modified BFGS method and its global convergence in nonconvex mini-mization, Journal of Computational and Applied Mathematics, 129(2001), pp. 15-35.

[30] D. Li and M. Fukushima, On the global convergence of the BFGS method for nonconvex uncon-strained optimization problems, SIAM Journal on Optimization, 11(2001), pp. 1054-1064.

[31] G. Li, C. Tang, and Z. Wei, New conjugacy condition and related new conjugate gradient methods for unconstrained optimization problems, Journal of Computational and Applied Mathemathics, 202(2007), pp. 532-539.

[32] G. H. Liu, J. Y. Han and D. F. Sun, Global convergence Analysis of the BFGS Algorithm withNonmonotone linesearch, Optimization, 34(1995), pp. 147-159.

[33] G. H. Liu, J. M. Peng, The convergence properties of a nonmonotonic algorithm, Journal of Com putational Mathematics, 1(1992), pp. 65-71.

[34] E. Polak, The conjugate gradient method in extreme problems, Computational Mathematics and Mathematical Physics, 9(1969), pp. 94-112.

[35] E. Polak and G. Ribie`re, Note sur la convergence de directions conjugees, Rev. Francaise informate Recherche Ope´ratinelle, 3(1969), pp. 35-43.

[36] M. J. D. Powell, Nonconvex minimization calculations and the conjugate gradient method, Lecture Notes in Mathematics, Vol. 1066, Spinger-Verlag, Berlin, 1984, pp. 122-141.

[37] M. J. D. Powell, Convergence properties of algorithm for nonlinear optimization, SIAM Review, 28(1986), pp. 487-500.

[38] P. L. Toint, An assessment of non-monotone line search techniques for unconstrained minimization problem, SIAM Journal on Computing, 17(1996), pp. 725-739.

[39] Z. Wei, G. Li and L. Qi, New quasi-Newton methods for unconstrained optimization problems, Applied Mathematics and Computation, 175(2006), pp. 1156-1188.

[40] 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, 29 (2004), pp. 315- 332.

[41] Z. Wei, G. Li, and L. Qi, New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems, Applied Mathematics and Computation, 179(2006), pp. 407-430.

[42] Z. Wei, S. Yao, and L. Liu, The convergence properties of some new conjugate gradient methods,Applied Mathematics and Computation, 183(2006), pp. 1341-1350.

[43] G. H. Yu, Nonlinear self-scaling conjugate gradient methods for large-scale optimization problems,thesis of Doctor's Degree, Sun Yat-Sen University, 2007.

[44] G. L. Yuan, Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems, Optimization Letters, 3(2009), pp. 11-21.

[45] G. L. Yuan, A conjugate gradient method for unconstrained optimization problems, International Journal of Mathematics and Mathematical Sciences, 2009(2009), pp. 1-14.

[46] G. L. Yuan and X. W. Lu, A modified PRP conjugate gradient method, Annals of Operations Research, 166(2009), pp. 73-90.

[47] G. L. Yuan, X. W. Lu, and Z. X. Wei, A conjugate gradient method with descent direction for unconstrained optimization, Journal of Computational and Applied Mathematics, 233(2009), pp. 519-530.

[48] Y. Yuan and W. Sun, Theory and methods of optimization, Science Press of China, Beijing, 1999.

[49] G. L. Yuan and Z. X. Wei, New line search methods for unconstrained optimization, Journal of the Korean Statistical Society, 38(2009), pp. 29-39.

[50] G. L. Yuan and Z. X. Wei, Convergence analysis of a modified BFGS method on convex minimiza tions, Computational Optimization and Applications, 47(2010), pp. 237-255.

[51] G. L. Yuan and Z. X. Wei, The superlinear convergence analysis of a nonmonotone BFGS algorithm on convex objective functions, Acta Mathematica Sinica, English Series, (1)24(2008), pp. 35-42.

[52] G. L. Yuan, Z. X. Wei, and Q. M. Zhao, A modified Polak-Ribi`re-Polyak conjugate gradient algo rithm for large-scale optimization problems, IIE Transactions, doi:10.1080/0740817X.2012.726757.

[53] G. L. Yuan and M. J. Zhang, A modified Hestens-Stiefel conjugate gradient algorithm for large-scale optimization, Numerical Functional Analysis and Optimization, 34(2013), pp. 914-937.

[54] J. Z. Zhang, N. Y. Deng, L. H. Chen, New quasi-Newton equation and related methods for uncon strained optimization, Journal Optimization Theory and Applications, 102(1999), pp. 147-167.

[55] H. C. Zhang and W. W. Hager, A nonmonotone line search technique and its application to unconstrained optimization, SIAM Journal on Optimization, 14(2004), pp. 1043-1056.

[56] L. Zhang, W. Zhou, and D. Li, A descent modified Polak-Ribi`re-Polyak conjugate method and its global convergence, IMA Journal on Numerical Analysis, 26(2006), pp. 629-649.

[57] J. L. Zhou and A. L. Tits, Nonmonotone line search for minimax problem, Journal Optimization Theory and Applications, 76(1993), pp. 455-476.

Appendix I. Numerical results of the paper.

problems	iteration number			nf+ng(the number of function and gradient)			cputime		
	HZ+	New-cg1	New-cg2	HZ+	New-cg1	New-cg2	HZ+	New-cg1	New-cg2
ARGLINA	1	1	1	5.00E+00	5	5	0.03599	0.03599	0.03499
ARGLINB	7	403	8	30	15644	207	0.04099	3.03754	0.07499
ARGLINC	227	6	9	1609	52	146	0.34195	0.04399	0.06099
ARWHEAD	9	8	8	31	27	27	0.07899	0.07899	0.07799
BDQRTIC	1217	635	832	4086	2480	4084	1.46678	0.99385	1.66575
BROWNAL	4	5	4	28	18	28	0.021	0.019	0.021
BROYDN7D	1444	1452	1456	4338	4362	4374	2.52062	2.53461	2.55561
BRYBND	31	31	31	98	98	98	0.09598	0.09698	0.09798
CHAINWOO	272	242	258	825	743	793	0.29095	0.26696	0.28196
CHNROSNB	245	279	286	737	841	860	0.003	0.004	0.003
COSINE	11	10	11	57	52	55	0.08099	0.07899	0.07999
CRAGGLVY	97	98	97	305	305	303	0.21397	0.21197	0.20897
CURLY10	64092	60002	67579	192346	180085	202815	101.55256	94.92457	108.89245
CURLY20	100367	96477	79307	301225	289520	238001	458.68527	282.859	228.2343
DECONVU	102	115	119	308	347	359	0.003	0.004	0.005
DIXMAANA	8	8	8	26	26	26	0.019	0.02	0.02
DIXMAANB	9	9	9	29	29	29	0.021	0.021	0.02
DIXMAANC	10	10	10	32	32	32	0.021	0.02	0.021
DIXMAAND	11	11	11	35	35	35	0.021	0.021	0.022
DIXMAANE	194	194	194	584	584	584	0.10098	0.10198	0.10198
DIXMAANF	147	147	147	443	443	443	0.08099	0.08099	0.08099
DIXMAANG	144	144	144	434	434	434	0.07999	0.07899	0.07899
DIXMAANH	140	140	140	422	422	422	0.07699	0.07799	0.07799
DIXMAANI	813	813	813	2441	2441	2441	0.37194	0.37494	0.37494
DIXMAANJ	137	137	137	413	413	413	0.07599	0.07699	0.07699
DIXMAANL	112	112	112	338	338	338	0.06499	0.06599	0.06599
DIXON3DQ	10000	10000	10000	30003	30003	30003	6.71398	6.87295	6.78597

DQDRTIC	7	7	7	23	23	23	0.04799	0.04899	0.04799
DQRTIC	32	32	32	98	98	98	0.02899	0.02799	0.02799
EDENSCH	29	30	31	89	92	95	0.03199	0.03399	0.03399
EG2	3	3	4	11	11	15	0.006	0.006	0.006
ENGVAL1	23	22	23	77	74	77	0.06399	0.06299	0.06299
ERRINROS	1069	1143	877	3460	3699	2869	0.011	0.012	0.01
EXTROSNB	3413	3059	2855	10572	9699	8968	0.38394	0.35794	0.33195
FLETCBV2	0	0	0	2	2	2	0.04899	0.04799	0.04899
FLETCHCR	6741	6669	6694	21343	21020	21048	1.10983	1.11483	1.10083
FMINSRF2	305	311	306	921	939	924	0.33595	0.34995	0.33795
FMINSURF	420	438	422	1262	1316	1268	0.46693	0.48792	0.47193
FREUROTH	65	63	52	214	316	204	0.14398	0.19497	0.14298
GENHUMPS	6718	6606	6704	20426	20038	20325	10.78636	10.20645	10.33443
GENROSE	1267	1250	1282	3862	3868	3984	0.10498	0.10498	0.10898
INDEF	1	1	1	105	105	105	0.11098	0.11098	0.10898
LIARWHD	21	25	19	80	97	68	0.05899	0.06399	0.05499
MANCINO	11	11	11	35	35	35	0.16097	0.16197	0.16297
MOREBV	32	32	32	99	99	99	0.05199	0.05199	0.05099
MSQRTALS	2443	2377	2377	7345	7147	7147	3.85441	3.74043	3.71343
MSQRTBLS	1907	1848	1840	5734	5560	5536	2.98955	2.91356	2.92056
NONCVXU2	7449	8819	7709	22349	26459	23131	7.93079	9.63354	8.34373
NONCVXUN	1058840	366321	940003	3176534	1098997	2820051	1115.76038	389.4398	997.67133
NONDIA	9	8	8	52	31	33	0.04499	0.03699	0.03799
NONDQUAR	2194	2236	1403	6658	6772	4275	0.84387	0.86287	0.55891
PENALTY1	43	43	43	170	170	170	0.01	0.011	0.01
PENALTY2	181	227	214	545	1031	918	0.021	0.03799	0.03399
POWELLSG	123	416	76	398	1354	239	0.06699	0.18097	0.04799
POWER	346	346	346	1040	1040	1040	0.17397	0.17897	0.17697
QUARTC	32	32	32	98	98	98	0.02899	0.02899	0.02899
SCHMVETT	35	36	34	109	112	106	0.11998	0.12298	0.11698

SENSORS	23	23	25	95	92	88	0.13598	0.12998	0.12298
SINQUAD	43	57	77	198	312	495	0.18697	0.23896	0.32495
SPARSINE	16571	11510	11650	49715	34532	34952	22.93851	15.90058	16.58948
SPARSQUR	20	20	20	62	62	62	0.12098	0.12098	0.12198
SPMSRTLS	186	188	190	574	580	586	0.27996	0.28396	0.28596
SROSENBR	12	11	11	42	38	38	0.02799	0.02699	0.02699
TESTQUAD	1623	1675	1675	4871	5027	5027	0.41094	0.43293	0.43193
TOINTGSS	3	3	3	11	11	11	0.02799	0.02799	0.02699
TOINTGOR	107	107	106	327	327	322	0.002	0.002	0.002
TOINTQOR	28	28	28	86	86	86	0.001	0.001	0.001
TQUARTIC	20	16	31	110	61	119	0.07599	0.06499	0.07599
TRIDIA	738	739	739	2216	2219	2219	0.26596	0.27096	0.26896
VARDIM	28	28	28	86	86	86	0.002	0.002	0.001
VAREIGVL	52	52	52	232	232	232	0.002	0.002	0.002
WOODS	155	124	238	577	512	797	0.11798	0.10798	0.15198