*J. Serv. Sci. & Management*, **2009, 2: 36-42**
Published Online March 2009 in SciRes (www.SciRP.org/journal/jssm)

# A Nonmonotone Line Search Method for Regression Analysis*

**Gonglin Yuan[1], Zengxin Wei[1]**

[1]College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, 530004, P. R. China.
Email: glyuan@gxu.edu.cn

Received January 7th, 2009; revised February 9th, 2009; accepted February 28th, 2009.

## ABSTRACT

*In this paper, we propose a nonmonotone line search combining with the search direction* (*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.) *for regression problems. The global convergence of the given method will be established under suitable conditions. Numerical results show that the presented algorithm is more competitive than the normal methods.*

***Keywords:*** *regression analysis, fitting method, optimization, nonmonotone, global convergence*

## 1. Introduction

It is well known that the regression analysis often arises in economies, finance, trade, law, meteorology, medicine, biology, chemistry, engineering, physics, education, history, sociology, psychology, and so on [1,2,3,4,5,6,7]. The classical regression model is defined by

$$Y=h(X_1, X_2, \ldots, X_p)+\varepsilon$$

where $Y$ is the response variable, Xi is predictor variable, $i$=1,2, …, $p$, $p>0$ is an integer constant, and $\varepsilon$ is the error. The function $h(X_1, X_2, \ldots, X_p)$ describes the relation between $Y$ and $X=(X_1, X_2, \ldots, X_p)$. If h is linear function, then we can get the following linear regression model

$$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p+\varepsilon \tag{1}$$

which is the most simple regression model, where $\beta_0$, $\beta_1$, …, $\beta_p$ are regression parameters. On the other hand, the regression model is called nonlinear regression. We all know that there are many nonlinear regression could be linearization [8,9,10,11,12,13]. Then many authors are devoted to the linear model [14,15,16,17,18,19]. Now we will concentrate on the linear model to discuss the following problems. One of the most important work of the regress analysis is to estimate the parameters $\beta=(\beta_0,\beta_1,\cdots,\beta_p)$.

The least squares method is an important fitting method to determined the parameters $\beta=(\beta_0,\beta_1,\cdots,\beta_p)$, which is defined by

$$\min_{\beta\in\Re^{p+1}} S(\beta)=\sum_{i=1}^{m}(h_i-(\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\cdots+\beta_pX_{ip}))^2 \tag{2}$$

*This work is supported by China NSF grands 10761001 and the Scientific Research Foundation of Guangxi University (Grant No. X081082).

where $h_i$ is the data valuation of the ith response variable, $X_{i1}$, $X_{i2}$, …, $X_{ip}$ are p data valuation of the ith predictor variable, and m is the number of the data. If the dimensionp and the number m is small, then we can obtain the parameters $\beta=(\beta_0,\beta_1,\cdots,\beta_p)$ from extreme value of calculus. From the definition (2), it is not difficult to see that this problem (2) is the same as the following unconstrained optimization problem

$$\min_{x\in\Re^n} f(x) \tag{3}$$

In this paper, we will concentrate on this problem (3) where $f:\ \Re^n\to\Re$ is continuously differentiable (linear or nonlinear). For regression problem (3), if the dimension n is large and the function f is complex, then the method of extreme value of calculus will fail. In order to solve this problem, numerical methods are often used, such as steepest descent method, Newton method, and Guass-Newton method [5,6,7]. Numerical method, i.e., the iterative method is to generates a sequence of points $\{x_k\}$ which will terminate or converge to a point $x^*$ in some sense. The line search method is one of the most effective numerical method, which is defined by

$$x_{k+1}=x_k+\alpha_k d_k, k=0,1,2,\cdots \tag{4}$$

where $\alpha_k$ is determined by a line search is the steplength, and $d_k$ which determines different line search methods [20,21,22,23,24,25,26,27] is a descent direction of f at $x_k$.

Due to its simplicity and its very low memory requirement, the conjugate gradient method is a powerful line search method for solving the large scale optimization problems. This method can avoid, like steepest de-

Copyright © 2009 SciRes 

scent method, the computation and storage of some matrices associated with the Hessian of objective functions. Conjugate gradient method has the form

$$d_{k+1} = \begin{cases} -g_{k+1} + \beta_k d_{k,} & if \ \ k \geq 1 \\ -g_{k+1}, & if \ \ k = 0 \end{cases} \tag{5}$$

where $g_k = \nabla f(x_k)$ is the gradient of f(x) at $x_k$, $\beta_k \in \Re$ is a scalar which determines the different conjugate gradient method [28,29,30,31,32,33,34,35,36,37]. 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. $\| . \|$ denotes the Euclidian norm of vectors. However, the following sufficiently des cent condition which is very important to insure the global convergence of the optimization problems

$$g_k^T dk \leq -c\|gk\|^2, \ for\ all\ k \geq 0 \text{ and some constant } c > 0 \tag{6}$$

is difficult to be satisfied by nonlinear conjugate gradient method, and this condition may be crucial for conjugate gradient methods [38]. At present, the global convergence of the PRP conjugate gradient method is still open when the weak Wolfe-Powell line search rule is used. Considering this case, Yuan and Wei [27] proposed a new direction defined by

$$d_{k+1} = \begin{cases} -g_{k+1} + \dfrac{\| g_{k+1} \|^2}{-g_{k+1}^T d_k} d_k, \ if\ g_{k+1}^T d_k \neq 0 \\ -g_{k+1} \qquad\qquad otherwise \end{cases} \tag{6}$$

where $d_0=-\nabla f_0=-g_0$. If $d_k^T g_{k+1} \neq 0$, it is easy to see that the search direction $d_k$ is the vector sum of the gradient $-g_k$ and the former search direction $d_{k-1}$, which is similar to conjugate gradient method. Otherwise, the steepest descent method is used as restart condition. Computational features should be effective. It is easy to see that the sufficiently descent condition (6) is true without carrying out any line search technique by this way. The global convergence has been established. Moreover, numerical results of the problems [39] and two regression analysis show that the given method is more competitive than the other similar methods [27].

Normally the steplength $\alpha_k$ is generated by the following weak Wolfe-Powell (WWP): Find a steplength $\alpha_k$ such that

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \sigma_1 \alpha_k g_k^T d_k \tag{8}$$

$$g_{k+1}^T d_k \geq \sigma_2 g_k^T d_k \tag{9}$$

where $0 < \sigma_1 < \sigma_2 < 1$. The monotone line search technique is often used to get the stepsize $\alpha_k$, however monotonicity may cause a series of very small steps if the contours of objective function are a family of curves with large curvature [40]. More recently, the nonmonotonic line search for solving unconstrained optimization is proposed by Grippo *et al*. in [40,41,42] and further studied by [43,44] etc. Grippo, Lamparillo, and Lucidi [40] proposed the following nonmonotone line search that they call it GLL line search. GLL line search: Select steplength $\alpha_k$ satisfying

$$f_{k+1} \leqslant \max_{0 \leq j \leq n(k)} f_{k-j} + \varepsilon_1 \alpha_k d_k^T g_k \tag{10}$$

$$g_{k+1}^T d_k \geq \max\{\varepsilon_2, 1-(\alpha_k \| d_k \|)^p\} g_k^T d_k \tag{11}$$

where $p \in (-\infty,1)$, $k$ = 0, 1, 2, …, $\varepsilon_1 \in (0,1), \varepsilon_2 \in (0,\frac{1}{2})$, $n(k) = \min\{H,k\}$, $H \geqslant 0$ is an integer constant. Combinng this line search and the normal BFGS formula, Han and Liu [45] established the global convergence of the convex objective function. Numerical results show that this method is more competitive to the normal BFGS method with WWP line search. Yuan and Wei [46] proved the superlinear convergence of the new nonmonotone BFGS algorithm.

Motivated by the above observations, we propose a nonmonotone method on the basic of Yuan and Wei [27] and Grippo, Lamparillo, and Lucidi [40]. The major contribution of this paper is an extension of the new direction in [27] to the nonmonotone line search scheme, and to concentrate on the regression analysis problems. Under suitable conditions, we establish the global convergence of the method. The numerical experiments of the proposed method on a set of problems indicate that it is interesting.

This paper is organized as follows. In the next section, the proposed algorithm is given. Under some reasonable conditions, the global convergence of the given method is established in Section 3. Numerical results and a conclusion are presented in Section 4 and in Section 5, respectively.

## 2. Algorithms

The proposed algorithm is given as follows.

Nonmonotone line search Algorithm (NLSA).

Step 0: Choose an initial point $x_0 \in \Re^n$, $0 < \varepsilon < 1$, $0 < \varepsilon_1 < \varepsilon_2 < 1$, $p \in (-\infty,1)$. an integer constant H>0. Set $d_0 = -\nabla f_0 = -g_0$, k :=0;

Step 1: If $\| g_k \|_2 \leqslant \varepsilon$, then stop; Otherwise go to step 2;

Step 2: Compute steplength $\alpha_k$ by Wolfe line search (10) and (11), let $x_{k+1} = x_k + \alpha_k d_k$.

Step 3: Calculate the search direction $d_{k+1}$ by (7).

Step 4: Set $k$: $=k+1$ and go to step 1.

Yuan and Wei [27] also presented two algorithms; here we stated them as follows. First another line search is given [47]: find a steplength $\alpha_k$ satisfying

$$f(x_k + \alpha_k d_k) \leq C_k + \sigma_1 \alpha_k g_k^T d_k \tag{12}$$

$$g_{k+1}^T d_k \geq \sigma_2 g_k^T d_k \tag{13}$$

Copyright © 2009 SciRes 

where $0 < \sigma_1 < \sigma_2 < 1$,

$$C_{k+1} = \frac{\mu_k Q_k C_k + f_{k+1}}{Q_{k+1}}, Q_{k+1} = \mu_k Q_k + 1,$$

$$C_0 = f_0, Q_0 = 1, \mu_k \in [\mu_{\min}, \mu_{\max}], 0 \le \mu_{\min} \le \mu_{\max} \le 1$$

**Algorithm 1 [27].**

Step 0: Choose an initial point $x_0 \in \Re^n$, $0 < \varepsilon < 1$, $0 < \sigma_1 < \sigma_2 < 1$. Set $d_0 = -\nabla f_0 = -g_0$, k :=0;

Step 1: If $\| g_k \|_2 \le \varepsilon$, then stop; Otherwise go to step 2;

Step 2: Compute steplength $\alpha_k$ by Wolfe line search (8) and (9), let $x_{k+1} = x_k + \alpha_k d_k$.

Step 3: Calculate the search direction $d_{k+1}$ by (7).

Step 4: Set $k$ :=$k$+1 and go to step 1.

**Algorithm 2 [27].**

Step 0: Choose an initial point $x_0 \in \Re^n$, $0 < \varepsilon < 1$, $0 < \mu < 1$, $0 < \sigma_1 < \sigma_2 < 1$. Set $C_0 = f_0, Q_0 = 1$, $d_0 = -\nabla f_0 = -g_0$, $k$:= 0;

Step 1: If $\| g_k \|_2 \le \varepsilon$, then stop; Otherwise go to step 2;

Step 2: Compute steplength $\alpha_k$ by the nonmonotone Wolfe line search (12) and (13), let $x_{k+1} = x_k + \alpha_k d_k$

Step 3: Calculate the search direction $d_{k+1}$ by (7).

Step 4: Let

$$Q_{k+1} = \mu Q_k + 1, C_{k+1} = \frac{\mu Q_k C_k + f_{k+1}}{Q_{k+1}} \tag{14}$$

Step 5: Set k: =k+1 and go to step 1.

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

## 3. Convergence Analysis

In order to establish the convergence of NLSA, the following assumptions are often needed [27,29,31,34,35,48].

Assumption 3.1: 1) f is bounded below on the bounded level set $\phi = \{x \in \Re^n : f(x) \leqslant f(x_0)\}$; 2) In $\phi$, f is differentiable and its gradient is Lipschitz continuous, namely, there exists a constants $L>0$ such that $\|g(x) - g(y)\| \leqslant L\|x - y\|$, for all $x, y \in \phi$.

In the following, we assume that $\|g_k\| \ne 0$ for all k, for otherwise a stationary point has been found. The following lemma shows that the search direction dk satisfies the sufficiently descent condition without any line search technique.

Lemma 3.1 (Lemma 3.1 in [27]) Consider (7). Then we have (6).

Based on Lemma 3.1 and Assumption 3.1, let us prove the global convergence theorem of NLSA.

Theorem 3.1 Let $\{\alpha_k, d_k, x_{k+1}, g_{k+1}\}$ be generated by the NLSA, and Assumption 3.1 holds. Then we have

$$\sum_{k=0}^{\infty} \left( \frac{g_k^T d_k}{\|d_k\|} \right)^2 < +\infty \tag{15}$$

and thus

$$\lim_{k \to \infty} \left( \frac{g_k^T d_k}{\| d_k \|} \right)^2 = 0 \tag{16}$$

Proof. Denote that

$$f(x_{h(k)}) = \max_{0 \le j \le n(k)} f(x_{k-j}), n(k) = \min\{H, k\}.$$

Using Lemma 3.1 and (10), we have

$$f_{k+1} \le \max_{0 \le j \le n(k)} f_{k-j} + \varepsilon_1 \alpha_k d_k^T g_k \le \max_{0 \le j \le n(k)} f_{k-j} = f(x_{h(k)})$$

Thus, we get

$$\begin{aligned} f(x_{h(k)}) &= \max_{0 \le j \le n(k)} f(x_{k-j}) \\ &\le \max\left\{ f(x_{h(k)}) = \max_{0 \le j \le n(k)} f(x_{k-1-j}), f_k \right\} \\ &= \max\left\{ f(x_{h(k-1)}), f(x_k) \right\} \\ &= f(x_{h(k-1)}), k = 1,2,..., \end{aligned} \tag{17}$$

i.e., the sequence $\{f(x_{h(k)}\}$ monotonically decreases. Since $f(x_{h(0)}) = f(x_0)$, we deduce that

$$f(x_k) \le f(x_{h(k-1)}) \le ... \le f(x_{h(0)}) = f_0$$

then $x_k \in \phi$. By Assumption 3.1: 1), we know that there exists a positive constant M such that

$$\| x \| \le M$$

Therefore,

$$\| \alpha_k d_k \| = \| x_{k+1} - x_k \| \le \| x_{k+1} \| + \| x_k \| \le 2M.$$

By (11), we have

$$\max\left\{\varepsilon_2, 1 - (\alpha_k \| d_k \|^p)\right\} \ge \max\left\{\varepsilon_2, 1 - (2M)\right\}^P$$

Let $\varepsilon_3 = \max\left\{\varepsilon_2, 1 - (2M)^P\right\} \in (0.1)$. Using (11) and Assumption 3.1: 2), we have

$$(\varepsilon_3 - 1) g_k^T d_k \le (g_{k+1} - g_k)^T d_k \le \| g_{k+1} - g_k \| \| d_k \| \le \alpha_k L \| d_k \|^2$$

Then we get

$$\alpha_k \ge \frac{(\varepsilon_3 - 1) g_k^T d_k}{L \| d_k \|^2} \tag{18}$$

By (10) and Lemma 3.1, we obtain

$$f_{k+1} \le f(x_{h(k)}) + \varepsilon_1 \alpha_k d_k^T g_k \le f(x_{h(k)}) - \frac{\varepsilon_1 (1 - \varepsilon_2)}{L} \left( \frac{d_k^T g_k}{\| d_k \|} \right)^2 \tag{19}$$

By Lemma 2.5 in [45], we conclude that from (19)

$$\sum_{k=0}^{\infty} \left( \frac{g_k^T d_k}{\| d_k \|} \right)^2 < +\infty \tag{20}$$

Therefore, (15) holds. (15) implies (16). The proof is complete.

Copyright © 2009 SciRes 

Remark. If there exists a constant $c_0>0$ such that $\| d_k \| \le c_0 \| g_k \|$ for all sufficiently large k. By (6) and (16), it is easy to obtain $\|g_k\| \to 0$ as $k\to\infty$.

## 4. Numerical Results

In this section, we report some numerical results with NLST, Algorithm 1, and Algorithm 2. All codes were written in MATLAB and run on PC with 2.60GHz CPU processor and 256MB memory and Windows XP operation system. The parameters and the rules are the same to those of [27], we state it as follows: $\sigma_1 = 0.1, \sigma_2 = 0.9, \mu = 10^{-4}, \varepsilon = 10^{-5}$. 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 fails. In order to avoid this case, the stepsize _k will be accepted if the searching number is more than twenty five in the line search. We will stop the program if the condition $\| \nabla f(\beta) \| 1e-5$ is satisfied. We also stop the program if the iteration number is more than one thousand, and the corresponding method is considered to be failed. In this experiment, the direction is defined by:

$$d_{k+1} = \begin{cases} -g_{k+1} + \dfrac{\| g_{k+1} \|^2}{-g_{k+1}^T d_k} d_k, & if\ g_{k+1}^T d_k < 1e-10 \\ -g_{k+1}, & otherwise \end{cases} \quad (21)$$

The parameters of the presented algorithm is chosen as: $\varepsilon_1 = 0.01, \varepsilon_2 = 0.1$, p=5, H=8.

In this section, we will test three practical problems to show the efficiency of the proposed algorithm, where Problem 1 and 2 can be seen from [27]. In Table 1 and 2, the initial points are the same to those of paper [27] and the results of Algorithm 1 and Algorithm 2 can also be seen from [27]. In order to show the efficiency of these algorithms, the residuals of sum of squares is defined by

$$SSE_p(\hat{\beta}) = \sum_{i=1}^{n} (y_i - \hat{y}_1)^2,$$

where $\hat{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + ... + \hat{\beta}_p X_{ip}$, i = 1, 2, …, n, and $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p$ are the parameters when the program is stopped or the solution is obtained from one way. Let

$$RMS_p(\hat{\beta}) = \frac{SSE_p(\hat{\beta})}{n-p}$$

where n is the number of terms in problems, and p is the number of parameters, if $RMS_p$ is smaller, then the corresponding method is better [49].

The columns of the tables 4−6 have the following meaning:

$\beta^*$: the approximate solution from the method of extreme value of calculus or some software. $\grave{\beta}$: the solution as the program is terminated. $\breve{\beta}$: the initial point. $\varepsilon_*$: the relative error between $RMS_p$ ($\beta^*$) and RMSp ($\grave{\beta}$) defined by $\varepsilon_* = \dfrac{RMS_p(\beta^*) - RMS_p(\beta)}{RMS_p(\beta^*)}$.

Problem 1. In the following table, there is data of some kind of commodity between year demand and price:

The statistical results indicate that the demand will possibly change though the price is inconvenient, and the demand will be possible invariably though the price changes. Overall, the demand will decrease with the increase of the price. Our objective is to find out the approximate function between the demand and the price, namely, we need to find the regression equation of d to the *p*.

It is not difficult to see that the price p and the demand d are linear relations. Denote the regression function by $\grave{d} = \beta_0 + \beta_1 p$, where $\beta_0$ and $\beta_1$ are the regression parameters.

Our work is to get $\beta_0$ and $\beta_1$. By least squares method, we need to solve the following problem

$$\min \sum_{i=0}^{n} (d_i - (\beta_0 + \beta_1 p_i))^2$$

and obtain $\beta_0$ and $\beta_1$, where n=10. Then the corresponding unconstrained optimization problem is defined by

$$\min_{\beta \in R^2} f(\beta) = \sum_{i=1}^{n} (d_i - \beta(1, p_i))^2 \quad (22)$$

Problem 2. In the following table, there is data of the age x and the average height H of a pine tree:

Similar to problem 1, it is easy to see that the age x and the average height H are parabola relations. Denote the regression function by $\hat{h} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, where $\beta_0$, $\beta_1$ and $\beta_2$ are the regression parameters. Using least squares method, we need to solve the following problem

$$\min \sum_{i=0}^{n} (h_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2))^2$$

and obtain $\beta_0$, $\beta_1$ and $\beta_2$, where n=10. Then the corresponding unconstrained optimization problem is defined by

$$\min_{\beta \in R^3} f(\beta) = \sum_{i=1}^{n} (h_i - \beta(1, x_i, x_i^2))^2 \quad (23)$$

It is well known that the above problems (22) and (24) can be solved by extreme value of calculus. Here we will solve these two problems by our methods and other two methods, respectively.

Problem 3. Supervisor Performance (Chapter 3 in [49]).

**Table 1. Demand and price**

| Price $p_i$($) Demand $d_i$ (500g) | 1<br>5 | 2<br>3.5 | 2<br>3 | 2.3<br>2.7 | 2.5<br>2.4 | 2.6<br>2.5 | 2.8<br>2 | 3<br>1.5 | 3.3<br>1.2 | 3.5<br>1.2 |
|---|---|---|---|---|---|---|---|---|---|---|

Copyright © 2009 SciRes 

**Table 2. Data of the age x and the average height H of a pine tree**

| $x_i$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|
| $h_i$ | 5.6 | 8 | 10.4 | 12.8 | 15.3 | 17.8 | 19.9 | 21.4 | 22.4 | 23.2 |

**Table 3. The data of appraisal to supervisor**

| line | Y | X1 | X2 | X3 | X4 | X5 | X6 |
|---|---|---|---|---|---|---|---|
| 1 | 43 | 51 | 30 | 39 | 61 | 92 | 45 |
| 2 | 63 | 64 | 51 | 54 | 63 | 73 | 47 |
| 3 | 71 | 70 | 68 | 69 | 76 | 86 | 48 |
| 4 | 61 | 63 | 45 | 47 | 54 | 84 | 35 |
| 5 | 81 | 78 | 56 | 66 | 71 | 83 | 47 |
| 6 | 43 | 55 | 49 | 44 | 54 | 49 | 34 |
| 7 | 58 | 67 | 42 | 56 | 66 | 68 | 35 |
| 8 | 71 | 75 | 50 | 55 | 70 | 66 | 41 |
| 9 | 72 | 82 | 72 | 67 | 71 | 83 | 31 |
| 10 | 67 | 61 | 45 | 47 | 62 | 80 | 41 |
| 11 | 64 | 53 | 53 | 58 | 58 | 67 | 34 |
| 12 | 67 | 60 | 47 | 39 | 59 | 74 | 41 |
| 13 | 69 | 62 | 57 | 42 | 55 | 63 | 25 |
| 14 | 68 | 83 | 83 | 45 | 59 | 77 | 35 |
| 15 | 77 | 77 | 54 | 72 | 79 | 77 | 46 |
| 16 | 81 | 90 | 50 | 72 | 60 | 54 | 36 |
| 17 | 74 | 85 | 64 | 69 | 79 | 79 | 63 |
| 18 | 65 | 60 | 65 | 75 | 55 | 80 | 60 |
| 19 | 65 | 70 | 46 | 57 | 75 | 85 | 46 |
| 20 | 50 | 58 | 68 | 54 | 64 | 78 | 52 |
| 21 | 50 | 40 | 33 | 34 | 43 | 64 | 33 |
| 22 | 64 | 61 | 52 | 62 | 66 | 80 | 41 |
| 23 | 53 | 66 | 52 | 50 | 63 | 80 | 37 |
| 24 | 40 | 37 | 42 | 58 | 50 | 57 | 49 |
| 25 | 63 | 54 | 42 | 48 | 66 | 75 | 33 |
| 26 | 66 | 77 | 66 | 63 | 88 | 76 | 72 |
| 27 | 78 | 75 | 58 | 74 | 80 | 78 | 49 |
| 28 | 48 | 57 | 44 | 45 | 51 | 83 | 38 |
| 29 | 85 | 85 | 71 | 71 | 77 | 74 | 55 |
| 30 | 82 | 82 | 39 | 59 | 64 | 78 | 39 |

where Y is overall appraisal to supervisor, $X_1$ denotes to processes employee's complaining, $X_2$ refer to do not permit the privilege, $X_3$ is the opportunity about study, $X_4$ is promoted based on the work achievement, $X_5$ refer to too nitpick to the bad performance, and $X_6$ is the speed of promoting to the better work. The above data can also be found at: http://www.ilr.cornell.edu/%7Ehadi/RABE3/Data/P054. txt.

Assume that the relation between Y and Xi (i=1, 2, …, 6) is linear [49], similar to Problem 1 and 2, the corresponding unconstrained optimization problem is defined by

$$\min_{\beta\in R^7} f(\beta)=\sum_{i=1}^{n}(h_i-\beta(1,x_{i1},x_{i2},...,x_{i6}))^2 \qquad (24)$$

where n = 30. The regression equation from one fitting way (see Chapter 3.8 in [49]) is given by

$$\hat{Y}=10.787+0.613X_1-0.073X_2+0.320X_3+0.081X_4+0.038X_5-0.217X_6$$

which means that $\beta^*$ =(10.787,0.613,−0.073,0.320, 0.081,0.038,−0.217). For Problem 3, the initial points are chosen as follows:

$\breve{\beta}_1$=(10, 0.1, −0.05, 1, 0.1, 2, −0.1); $\breve{\beta}_2$=(10, −0.1, 0.05, −1, −0.1, −2, 0.1);

$\breve{\beta}_3$=(10.1, −0.01, 0.5, −0.2, −0.01, −0.2, 4); $\breve{\beta}_4$=(10.8, −100, 20, −70, −50, −40, 60);

$\breve{\beta}_5$= (9, 0.01, −0.5, 1, 0.01, 2, −0.01); $\breve{\beta}_6$= (11, 0.01, −0.5, 1, 0.01, 2, −0.01).

These numerical results of Table 4-6 indicate that proposed algorithm is more competitive than those of Algorithm 1 and 2, and the initial points do not influence the results obviously about these three methods. Moreover, the numerical results of NLSA, Algorithm 1, and Algorithm 2 are better than those of these methods from extreme value of calculus or some software. Then we can conclude that the numerical method will outperform the method of extreme value of calculus in some sense, and some software for regression analysis could be further improved in the future. Overall, the direction defined by (7) is notable.

**Table 4. Test results for Problem 1**

| $\beta^*$=(6.5–1.6) | $\breve{\beta}$ | $\grave{\beta}$ | RMSp ($\grave{\beta}$) | RMSp($\beta^*$) | $\varepsilon_*$ |
|---|---|---|---|---|---|
| Algorithm 1 | (1, −0.01) | (6.438301, −1.575289) | 0.039736 | 0.040100 | 0.908% |
| | (−10,0.04) | (6.438280, −1.575313) | 0.039736 | 0.040100 | 0.908% |
| | (−2, −1.0) | (6.438285, −1.575314) | 0.039736 | 0.040100 | 0.908% |
| | (15,15) | (6.438287, −1.575316) | 0.039736 | 0.040100 | 0.908% |
| Algorithm 2 | (1, −0.01) | (6.438301, −1.575289) | 0.039736 | 0.040100 | 0.908% |
| | (−10,0.04) | (6.438280, −1.575313) | 0.039736 | 0.040100 | 0.908% |
| | (−2,-1.0) | (6.438285, −1.575314) | 0.039736 | 0.040100 | 0.908% |
| | (15,15) | (6.438287, −1.575316) | 0.039736 | 0.040100 | 0.908% |
| NLSA | (1, −0.01) | (6.438280, −1.575312) | 0.039736 | 0.040100 | 0.908% |
| | −10,0.04) | (6.438292, −1.575317) | 0.039736 | 0.040100 | 0.908% |
| | (−2,-1.0) | (6.438291, −1.575316) | 0.039736 | 0.040100 | 0.908% |
| | (15,15) | (6.438280, −1.575312) | 0.039736 | 0.040100 | 0.908% |

Copyright © 2009 SciRes 

**Table 5. Test results for Problem 2**

| $\beta^*$=(−1.33, 3.46, −0.11) | $\beta$ | $\beta$ | RMSp ($\beta$) | RMSp ($\beta^*$) | $\varepsilon_*$ |
|---|---|---|---|---|---|
| Algorithm 1 | (−1.1,3.0, −0.5) | (−1.296574, 3.450247, −0.107896) | 0.171774 | 0.183900 | 6.5938% |
| | (−1.2,3.2, −0.3) | (−1.328742, 3.460876, −0.108650) | 0.171712 | 0.183900 | 6.6273% |
| | (−0.003,7.0, −0.8) | (−1.328504, 3.460798, −0.108646) | 0.171713 | 0.183900 | 6.6272% |
| | (−0.001,7.0, −0.5) | (−1.321726, 3.458558, −0.108483) | 0.171717 | 0.183900 | 6.6248% |
| Algorithm 2 | (−1.1,3.0, −0.5) | (−1.296574, 3.450247, −0.107896) | 0.171774 | 0.183900 | 6.5938% |
| | (−1.2,3.2, −0.3) | (−1.328742, 3.460876, −0.108650) | 0.171712 | 0.183900 | 6.6273% |
| | (−0.003,7.0, −0.8) | (−1.328504, 3.460798, −0.108646) | 0.171713 | 0.183900 | 6.6272% |
| | (−0.001,7.0, −0.5) | (−1.321726, 3.458558, −0.108483) | 0.171717 | 0.183900 | 6.6248% |
| NLSA | (−1.1,3.0, −0.5) | (−1.331296, 3.461720, −0.108711) | 0.171712 | 0.183900 | 6.6274% |
| | (−1.2,3.2, −0.3) | (−1.331232, 3.461699, −0.108709) | 0.171712 | 0.183900 | 6.6274% |
| | (−0.003,7.0, −0.8) | (−1.331140, 3.461669, −0.108707) | 0.171712 | 0.183900 | 6.6274% |
| | (−0.001,7.0, −0.5) | (−1.202673, 3.422106, −0.106011) | 0.172583 | 0.183900 | 6.1539% |

**Table 6. Test results for Problem 2**

| $\beta^*$ | $\beta$ | $\beta$ | RMSp($\beta$) | RMSp($\beta^*$) | $\varepsilon_*$ |
|---|---|---|---|---|---|
| Algorithm 1 | $\bar{\beta}_1$ | (10.011713, 0.502264, -0.002329, 0.361596, 0.061871, 0.152295, -0.353686) | 85.261440 | 89.584291 | 4.8255% |
| | $\bar{\beta}_2$ | (10.124457, 0.502394, -0.002598, 0.361313, 0.061446, 0.151381, -0.353527) | 85.235105 | 89.584291 | 4.8549% |
| | $\bar{\beta}_3$ | (10.294617, 0.502056, -0.002462, 0.360523, 0.062746, 0.149161, -0.354270) | 85.196215 | 89.584291 | 4.8983% |
| | $\bar{\beta}_4$ | (11.404702, 0.501820, -0.004943, 0.357265, 0.060921, 0.140326, -0.354036) | 84.963796 | 89.584291 | 5.1577% |
| | $\bar{\beta}_5$ | (9.542516, 0.503279, -0.001805, 0.362715, 0.061217, 0.156318, -0.352638) | 85.375457 | 89.584291 | 4.6982% |
| | $\bar{\beta}_6$ | (11.071364, 0.501290, -0.004085, 0.358312, 0.062185, 0.143081, -0.354614) | 85.029566 | 89.584291 | 5.0843% |
| Algorithm 2 | $\bar{\beta}_1$ | (10.011713, 0.502264, -0.002329, 0.361596, 0.061871, 0.152295, -0.353686) | 85.261440 | 89.584291 | 4.8255% |
| | $\bar{\beta}_2$ | (10.166214, 0.502293, -0.002549, 0.360902, 0.062002, 0.151044, -0.354147) | 85.225461 | 89.584291 | 4.8656% |
| | $\bar{\beta}_3$ | (10.639778, 0.502423, -0.003742, 0.360018, 0.060167, 0.147253, -0.353327) | 85.119812 | 89.584291 | 4.9836% |
| | $\bar{\beta}_4$ | (11.404239, 0.501827, -0.004935, 0.357227, 0.060988, 0.140322, -0.354037) | 84.963893 | 89.584291 | 5.1576% |
| | $\bar{\beta}_5$ | (11.404239, 0.501827, -0.004935, 0.357227, 0.060988, 0.140322, -0.354037) | 85.506424 | 89.584291 | 4.5520% |
| | $\bar{\beta}_6$ | (11.032035, 0.501940, -0.004251, 0.358407, 0.061171, 0.143518, -0.353940) | 85.037491 | 89.584291 | 4.5520% |
| NLSA | $\bar{\beta}_1$ | (10.326165, 0.502177, -0.002900, 0.360625, 0.061701, 0.149611, -0.353760) | 85.189017 | 89.584291 | 4.9063% |
| | $\bar{\beta}_2$ | (10.042910, 0.501267, -0.001983, 0.359836, 0.065677, 0.151241, -0.354909) | 85.254692 | 89.584291 | 4.8330% |
| | $\bar{\beta}_3$ | (10.525637, 0.502094, -0.003292, 0.359987, 0.061542, 0.147873, -0.353823) | 85.144572 | 89.584291 | 4.9559% |
| | $\bar{\beta}_4$ | (11.431772, 0.501805, -0.005001, 0.357160, 0.060909, 0.140080, -0.354047) | 84.958622 | 89.584291 | 5.1635% |
| | $\bar{\beta}_5$ | (9.653770, 0.502364, -0.001653, 0.362701, 0.062144, 0.155364, -0.353611) | 85.347711 | 89.584291 | 4.7292% |
| | $\bar{\beta}_6$ | (11.504977, 0.501791, -0.005132, 0.356938, 0.060866, 0.139459, -0.354060) | 84.944709 | 89.584291 | 5.1790% |

## 5. Conclusions

The major contribution of this paper is an extension of the direction (7) to a nonmonotone line search technique (GLL line search). The presented method possess global convergence and the numerical results show that the given algorithm is successful for the test problems. These test numerical results further show that the direction defined by (7) is notable. We hope the method can be a further topic for the regression analysis.

For further research, we should study other line search methods for regression analysis.

Moreover, more numerical experiments for large practical problems about regression analysis should be done in the future.

## REFERENCES


[1] D. M. Bates and D. G. Watts, “Nonlinear regression analysis and its applications,” New York: John Wiley & Sons, 1988.

[2] S. Chatterjee and M. Machler, “Robust regression: A weighted least squares approach, communications in statistics,” Theorey and Methods, 26, pp. 1381−1394, 1997.

[3] R. Christensen, “Analysis of variance, design and regression: Applied statistical methods,” New York: Chapman and Hall, 1996.

[4] N. R. Draper and H. Smith, “Applied regression analysis,” 3rd ed., New York: John Wiley & Sons, 1998.

[5] F. A. Graybill and H. K. Iyer, “Regression analysis: Concepts and applications, Belmont,” CA: Duxbury Press, 1994.

[6] R. F. Gunst and R. L. Mason, “Regression analysis and its application: A data-Oriented approach,” New York: Marcel Dekker, 1980.

[7] R. H. Myers, “Classical and modern regression with applications,” 2nd edition, Boston: PWS-KENT Publishing Company, 1990.

[8] R. C. Rao, “Linear statistical inference and its applications,”New York: John Wiley & Sons, 1973.

[9] D. A. Ratkowsky, “Nonlinear regression modeling: A unified practical approach,” New York: Marcel Dekker, 1983.


Copyright © 2009 SciRes 

[10] D. A. Ratkowsky, “Handbook of nonlinear regression modeling,” New York: Marcel Dekker, 1990.

[11] A. C. Rencher, “Methods of multivariate analysis,” New York: John Wiley & Sons, 1995.

[12] G. A. F. Seber and C. J. Wild, “Nonlinear regression,” New York: John Wiley & Sons, 1989.

[13] A. Sen and M. Srivastava, “Regression analysis: Theory, methods, and applications,” New York: Springer-Verlag, 1990.

[14] J. Fox, “Linear statistical models and related methods,” New York: John Wiley & Sons, 1984.

[15] S. Haberman and A. E. Renshaw, “Generalized linear models and actuarial science,” The Statistician, 45, pp. 407−436, 1996.

[16] S. Haberman and A. E. Renshaw, “Generalized linear models and excess mortality from peptic ulcers,” Insurance: Mathematics and Economics, 9, pp. 147−154, 1990.

[17] R. R. Hocking, “The analysis and selection of variables in linear regression,” Biometrics, 32, pp. 1−49, 1976.

[18] P. McCullagh and J. A. Nelder, “Generalized linear models,” London: Chapman and Hall, 1989.

[19] J. A. Nelder and R. J. Verral, “Credibility theory and generalized linear models,” ASTIN Bulletin, 27, pp. 71−82, 1997.

[20] M. Raydan, “The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem,”SIAM Journal of Optimization, 7, pp. 26−33, 1997.

[21] J. Schropp, “A note on minimization problems and multistep methods,” Numerical Mathematics, 78, pp. 87−101, 1997.

[22] J. Schropp, “One-step and multistep procedures for constrained minimization problems,” IMA Journal of Numerical Analysis, 20, pp. 135−152, 2000.

[23] D. J. Van. Wyk, “Differential optimization techniques,” Appl. Math. Model, 8, pp. 419−424, 1984.

[24] M. N. Vrahatis, G. S. Androulakis, J. N. Lambrinos, and G. D. Magolas, “A class of gradient unconstrained minimization algorithms with adaptive stepsize,” Journal of Computational and Applied Mathematics, 114, pp. 367−386, 2000.

[25] G. L. Yuan and X. W. Lu, “A new line search method with trust region for unconstrained optimization,” Communications on Applied Nonlinear Analysis, Vol. 15, No. 1, pp. 35−49, 2008.

[26] G. Yuan, X. Lu, and Z. Wei, “New two-point stepsize gradient methods for solving unconstrained optimization problems,” Natural Science Journal of Xiangtan University, (1)29, pp. 13−15, 2007.

[27] G. L. Yuan and Z. X. Wei, “New line search methods for unconstrained optimization,” Journal of the Korean Statistical Society, 38, pp. 29−39, 2009.

[28] Y. Dai, “A nonmonotone conjugate gradient algorithm for unconstrained optimization,” Journal of Systems Science and Complexity, 15, pp. 139−145, 2002.

[29] Y. Dai and Y. Yuan, “A nonlinear conjugate gradient with a strong global convergence properties,” SIAM Journal of Optimization, 10, pp. 177−182, 2000.

[30] R. Fletcher, “Practical Method of Optimization,” Vol 1: Unconstrained Optimization, 2nd edition, Wiley, New York, 1997.

[31] R. Fletcher and C. Reeves, “Function minimization by conjugate gradients,” The Computer Journal, 7, pp, 149−154, 1964.

[32] Y. Liu and C. Storey, “Effcient generalized conjugate gradient algorithms, part 1: theory,” Journal of Optimization Theory and Application, 69, pp. 17−41, 1992.

[33] E. Polak and G. Ribiere, “Note sur la convergence de directions conjugees,” Rev. Francaise informat Recherche Operatinelle, 3e Annee, 16, pp. 35−43, 1969.

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

[35] Z. Wei, S. Yao, and L. Liu, “The convergence properties of some new conjugate gradient methods,” Applied Mathematics and Computation, 183, pp. 1341−1350, 2006.

[36] G. L. Yuan, “Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems,” Optimization Letters, DOI: 10. 1007/s11590−008−0086−5, 2008.

[37] G. L. Yuan and X. W. Lu, “A modified PRP conjugate gradient method,” Annals of Operations Research, 166, pp. 73−90, 2009.

[38] J. C. Gibert and J. Nocedal, “Global convergence properties of conjugate gradient methods for optimization,” SIAM Journal of Optimization, 2, pp. 21−42, 1992.

[39] J. J. Mor´e, B. S. Garbow, and K. E. Hillstrome, “Testing unconstrained optimization software,” ACM Transactions Math. Software, 7, pp. 17−41, 1981.

[40] L. Grippo, F. Lamparillo, and S. Lucidi, “A nonmonotone line search technique for Newton’s method,” SIAM Journal of Numerical Analysis, 23, pp. 707−716, 1986.

[41] L. Grippo, F. Lamparillo, and S. Lucidi, “A truncate Newton method with nonmonotone line search for unconstrained optimization,” Journal of Optimization Theory and Applications, 60, pp. 401−419, 1989.

[42] L. Grippo, F. Lamparillo, and S. Lucidi, “A class of nonmonotone stabilization methods in unconstrained optimization,” Numerical Mathematics, 59, pp. 779−805, 1991.

[43] G. H. Liu, J. Y. Han, and D. F. Sun, “Global convergence analysis of the BFGS algorithm with nonmonotone linesearch,” Optimization, Vol. 34, pp. 147−159, 1995.

[44] G. H. Liu, J. M. Peng, The convergence properties of a nonmonotonic algorithm,” Journal of Computational Mathematics, 1, pp. 65−71, 1992.

[45] 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, pp. 277−289, 1997.

[46] 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, Vol. 24, No. 1, pp. 35−42, 2008.

[47] H. C. Zhang and W. W. Hager, “A nonmonotone line search technique and its application to unconstrained optimization,” SIAM Journal of Optimization, Vol. 14, No. 4, pp. 1043−1056, 2004.

[48] M. R. Hestenes and E. Stiefel, “Method of conjugate gradient for solving linear equations,” J, Res. Nat. Bur. Stand., 49, pp. 409−436, 1952.

[49] S. Chatterjee, A. S. Hadi, and B. Price, “Regression analysis by example,” 3rd Edition, John Wiley & Sons, 2000.


**(Edited by Vivian and Ann)**

Copyright © 2009 SciRes 