*Applied Mathematics*, 2011, 2, 39-46
doi:10.4236/am.2011.21005 Published Online January 2011 (http://www.SciRP.org/journal/am)

# A Gauss-Newton-Based Broyden's Class Algorithm for Parameters of Regression Analysis

**Xiangrong Li, Xupei Zhao**
*Department of Mathematics and Information Science*, *Guangxi University*, *Nanning*, *China*
*E-mail*: *xrli68@163.com*
*Received October* 14, 2010; *revised November* 11, 2010; *November* 13, 2010

## Abstract

In this paper, a Gauss-Newton-based Broyden's class method for parameters of regression problems is presented. The global convergence of this given method will be established under suitable conditions. Numerical results show that the proposed method is interesting.

**Keywords:** Global Convergence, Broyden's Class, Regression Analysis, Nonlinear Equations, Gauss-Newton

## 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 (see [1-7]). The classical regression model is defined by

$$Y = h\left(X_1, X_2, \cdots, X_p\right) + \varepsilon,$$

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

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \tag{1.1}$$

which is the most simple regression model, where $\beta_0, \beta_1, \cdots, \beta_p$ 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 [8-13].Then many authors are devoted to the linear model [14-19]. 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 $\beta = \left(\beta_0, \beta_1, \cdots, \beta_p\right)$. The least squares method is an important fitting method to determined the parameters $\beta = \left(\beta_0, \beta_1, \cdots, \beta_p\right)$, which is defined by

$$\min_{\beta \in \Re^{p+1}} S(\beta) = \sum_{i=1}^{m} \left(h_i - \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}\right)^2, \tag{1.2}$$

where $h_i$ is the data valuation of the $i$ th response variable, $X_{i1}, X_{i2}, \cdots, X_{ip}$ are $p$ data valuation of the $i$ th predictor variable, and $m$ is the number of the data. If the dimension $p$ and the number $m$ is small, then we can obtain the parameters $\beta = \left(\beta_0, \beta_1, \cdots, \beta_p\right)$ from extreme value of calculus. From the definition of (1.2), it is not difficult to see that this problem (1.2) is the same as the following unconstrained optimization problem

$$\min_{x \in \Re^n} f(x) \tag{1.3}$$

For regression problem (1.3), if the dimension $n$ is large and the function $f$ is complex, then it is difficult to solve this problem by the method of extreme value of calculus. In order to solve this problem, numerical methods are often used, such as steepest descent method, Newton method, and Guass-Newton method (see [5-7] *et al.*). Moreover many statical softwares are from this idea. 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

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

where $\alpha_k$ that is determined by a line search is the steplength, and $d_k$ which determines different line search methods [20-30] is a descent direction of $f$ at $x_k$. We

This work is supported by China NSF grands 10761001 and Guangxi SF grands 0991028.

 

give a line search method for regression problem and get good results (see [31] in detail).

In order to solve the problem (1.3), one main goal is to find some point $x$ such that

$$g(x) = 0, x \in \Re^n \quad (1.4)$$

where $g(x) = \nabla f(x)$ is the gradient of $f(x)$ In this paper, we will concentrate on this equations problem (1.4) where $g: \Re^n \to \Re^2$ is continuously differentiable (linear or nonlinear). Assume that the Jacobian $\nabla g(x)$ of $g$ is symmetric for all $x \in \Re^n$. Let $\theta$ be the norm function defined by $\theta(x) = \frac{1}{2}\|g(x)\|^2$.Then the nonlinear equations problem (1.4) is equivalent to the following global optimization problem

$$\min \theta(x), x \in \Re^n . \quad (1.5)$$

Similar to (1.3), the following iterative formula is often used to solve the problem (1.4) or (1.5)

$$x_{k+1} = x_k + \alpha_k d_k , \quad (1.6)$$

where $d_k$ is a search direction, $\alpha_k$ is a steplength along $d_k$ and $x_k$ is the $k$ th iterative point. For (1.4), Griewank [32] first established a global convergence theorem for quasi-Newton method with a suitable line search. One nonmonotone backtracking inexact quasi-Newton algorithm [33] and the trust region algorithms [34,35] were presented. A Gauss-Newton-based BFGS (Broyden, Fletcher, Goldfar, and Shanno, 1970) method is proposed by Li and Fukushima [36] for solving symmetric nonlinear equations. Inspired by their ideas, Wei [37] and Yuan [38,39] made a further study. Recently, Yuan and Lu [40-45] got some new methods for symmetric nonlinear equations.

The authors [36] only discussed that the updated matrices were generated by the BFGS formula. Whether the updated matrices could be produced by the more extensive Broyden's class? This paper gives a positive answer, moreover, the presented method is used to regression analysis. The major contribution of this paper is an extension of the method in [36] to Broyden’s class, moreover, to solving the regression problems. Numerical results of practically statistical problems show that this given method is effective. Throughout this paper, these notations are used: $\|\cdot\|$ is the Euclidean norm, $g(x_k)$ and $g(x_{k+1})$ are replaced by $g_k$ and $g_{k+1}$, respectively.

In the next section, the method of Li and Fukushima [36] is stated. Our algorithm is proposed in Section 3. Under some reasonable conditions, the global convergence of the given algorithm is established in Section 4. In the Section 5, numerical results are reported. In the last section, a conclusion is stated.

## 2. A Gauss-Newton-Based BFGS Method [36]

Li and Fukushima [36] proposed a new BFGS update formula defined by:

$$B'_{k+1} = B'_k - \frac{B'_k s_k s_k^T B'_k}{s_k^T B' s_k} + \frac{\delta_k \delta_k^T}{\delta_k^T s_k}, \quad (2.1)$$

Where $s_k = x_{k+1} - x_k, y_k = g_{k+1} - g_k$, $\delta_k = g(x_k + y_k) - g_k, x_{k+1}$ is the next iteration, $g_k = g(x_k)$, $g_{k+1} = g(x_{k+1})$, and $B'_0$ is an initial symmetric positive definite matrix. By the secant equation $B'_{K+1} s_k = \delta_k$ and $\nabla g_k$ is symmetric, they had approximately

$$B'_{k+1} s_k \approx \nabla g_{k+1} y_k \approx \nabla g_{k+1}^T \nabla g_{k+1} s_k ,$$

which implies that $B'_{k+1}$ approximates $\nabla g_{k+1}^T \nabla g_{k+1}$ along direction $s_k$ By solving the following linear equation to get the search direction $d_k$.

$$B_k d_k + \frac{g(x_k + \alpha_{k-1} g_k) - g_k}{\alpha_{k-1}} = 0. \quad (2.2)$$

If $\|\alpha_{k-1} g_k\|$ is sufficiently small and $B_k$ is positive definite, then they have the following approximate relation

$$B_k d_k = -\frac{g(x_k + \alpha_{k-1} g_k) - g_k}{\alpha_{k-1}} \approx -\nabla g_k g_k .$$

Therefore,

$$d_k \approx -B_k^{-1} \nabla g_k g_k \approx -\left(\nabla g_k^T \nabla g_k\right)^{-1} \nabla g_k g_k .$$

So, the solution of (2.2) is an approximate Gauss-Newton direction. Then the methods (2.1) and (2.2) are called Gauss-Newton-based BFGS method. In order to get the steplength $\alpha = \alpha_k$ by means of a backtracking process, a new line search technique is defined by

$$\begin{aligned} &\|g(x_k + \alpha d_k)\|^2 - \|g_k\|^2 \\ &\le -\sigma_1 \|\alpha g_k\|^2 - \sigma_2 \|\alpha d_k\|^2 + \varepsilon_k \|g_k\|^2 , \end{aligned} \quad (2.3)$$

where $\sigma_1, \sigma_2 > 0$ are constants, and the positive sequence $\{\varepsilon_k\}$ such that

$$\sum_{k=0}^{\infty} \varepsilon_k < \infty . \quad (2.4)$$

Li and Fukushima [36] only discussed that the updated matrices were generated by the BFGS formula. In this paper, we will prove that the updated matrices could be produced by the more extensive Broyden's class. Moreover, the presented method is used to regression analysis (1.3) Numerical results show that the given method is promising.

## 3. Algorithm

Now we give our algorithm as follows.

**Algorithm 1** (*Gauss-Newton-based Broyden's Class Algorithm*)

Step 0: Choose an initial point $x_0 \in R^n$ ,an initial symmetric positive definite matrix $B_0 \in R^{n\times n}$ , a positive sequence $\{\varepsilon_k\}$ satisfying (2.4), and constants $r \in (0,1), \sigma_1, \sigma_2 > 0, \alpha_{-1} > 0$ , let: $k = 0$ .

Step 1: Stop if $\|g_k\| = 0$. Otherwise solve the linear (2.2) to get the search direction $d_k$ .

Step 2: Let $i_k$ be the smallest nonnegative integer $i$ such that Equation (2.3) holds for $\alpha = r^i$ . Let $\alpha_k = r^i$ .

Step 3: Let the next iterative be $x_{k+1} = x_k + \alpha_k d_k$ .

Step 4: Put $s_k = x_{k+1} - x_k = \alpha_k d_k, \delta_k = g_{k+1} - g_k$ and $y_k = g(x_k + \delta_k) - g(x_k)$. If $s_k^T y_k > 0$ and

$$\phi_k > \phi_k^c = \frac{1}{1-u_k}, u_k = \frac{\left(s_k^T B_k s_k\right)\left(y_k^T H_k y_k\right)}{\left(s_k^T y_k\right)^2}, H_k = B_k^{-1}, \tag{3.1}$$

then update $B_k$ by the Broyden's class formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} + \phi_k s_k^T B_k s_k v_k v_k^T, \quad k = 0,1,2,\cdots, \tag{3.2}$$

Where $v_k = \frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k}$ Otherwise let $B_{k+1} = B_k$ .

Step 5: Let $k := k+1$ Go to step 1.

## 4. Global Convergence

In this paper, we will establish the global convergence of Algorithm 1 under the condition about $\phi_k$ such that

$$\varphi_k \in \left[(1-v)\varphi_k^c, 1-\ell\right], \ell, v \in (0,1) \tag{4.1}$$

Let $\Omega$ be the level set defined by $\Omega = \left\{ x \mid \|g(x)\| \le e^{\frac{\tau}{2}} \|g(x_0)\| \right\}$, where $\varepsilon$ is a positive constant such that $\sum_{k=0}^{\infty} \varepsilon_k \le \varepsilon$ .

Similar to [33,36-39], the following Assumptions are needed.

**Assumption A** 1) $g$ is continuously differentiable on an open convex set $\Omega_1$ containing $\Omega$ .

2) The Jaconbian of $g$ is symmetric, bounded, and uniformly nonsingular on $\Omega_1$ , i.e., there exist positive constants $M \ge m > 0$ such that

$$\|\nabla g(x)\| \le M \quad \forall x \in \Omega_1 \tag{4.2}$$

and

$$m\|d\| \le \|\nabla g(x) d\|, \ \forall x \in \Omega_1, d \in R^n. \tag{4.3}$$

Assumption A 2) implies that

$$m\|d\| \le \|\nabla g(x) d\| \le M\|d\|, \forall x \in \Omega_1, d \in R^n, \tag{4.4}$$

$$m\|x-y\| \le \|g(x) - g(y)\| \le M\|x-y\|, \forall x, y \in \Omega_1. \tag{4.5}$$

By Assumption A, similar to Lemma 2.2 in [36], it is not difficult to get the following lemma. So we only state it as follows but omit the proof.

**Lemma 4.2** Let Assumption $A$ be satisfied. Consider Equation (2.3), if $s_k \to 0$ , then there is a constant $m_1 > 0$ such that for all $k$ sufficiently large

$$y_k^T s_k \ge m_1 \|s_k\|^2. \tag{4.6}$$

Denote

$$\begin{aligned} q_k &= \frac{g(x_k + \alpha_{k-1} g_k) - g_k}{\alpha_{k-1}} \\ &= \int_0^1 \nabla g(x_k + \tau\alpha_{k-1} g_k) d\tau g_k = T_k g_k, \end{aligned} \tag{4.7}$$

where $T_k = \int_0^1 \nabla g(x_k + \tau\alpha_{k-1} g_k) d\tau$ . Consider Equation (2.2), then we have

$$B_k d_k + q_k = B_k d_k + T_k g_k = 0. \tag{4.8}$$

**Lemma 4.3** *Let Assumption $A$ be satisfied. Then we have*

$$\lim_{k\to\infty} \frac{\left(s_k^T \alpha_k q_k\right)^2}{s_k^T y_k} = 0. \tag{4.9}$$

**Proof.** Consider the line search (2.3), by Lemma 4.1 and Equation (2.4), we can get the following inequalities immediately

$$\sum_{k=0}^{\infty} \|\alpha_k g_k\|^2 < \infty, \sum_{k=0}^{\infty} \|\alpha_k d_k\|^2 < \infty. \tag{4.10}$$

By Equations (4.5)-(4.7), we have

$$0 \le \frac{\left(\alpha_k s_k^T q_k\right)^2}{s_k^T y_k} \le \frac{1}{m_1} \alpha_k^2 \|q_k\|^2 \le \frac{M^2}{m_1} \|\alpha_k g_k\|^2,$$

By Equation (4.10), we obtain Equation (4.9). The proof is complete.□

**Lemma 4.4**

$$\det(B_{K+1}) = \det(B_k)\left(y_k^T s_k / s_k^T B_k s_k\right)\left\{1 + \varphi_k (u_k - 1)\right\}$$

where $\det(B_k)$ denotes the determinant of $B_k$ .

**Proof.** Omitted. For the proof can be seen from [21].□

Let us denote

$$\cos\theta_k = \frac{s_k^T B_k s_k}{\|B_k s_k\| . \|s_k\|}. \tag{4.11}$$

The proof of the following lemma is motivated by the methods in [46,47].

**Lemma 4.5** *Let Assumption A hold and* $\{\alpha_k, d_k, x_k, g_k\}$ *be generated by Algorithm* 1. *Then there exist a positive integer* $k'$ *and positive constants* $\beta_1, \beta_2, \beta_3 > 0$ *such that, for any* $t_0 \in (0,1)$ *and* $k \ge k'$ *the relations*

$$\cos\theta_i \ge \beta_1, \beta_2 \le s_i^T B_i s_i \le \beta_3, \beta_2 \|s_i\| \le \|B_i s_i\| \le \frac{\beta_3}{\beta_1} \quad (4.12)$$

*hold for at least* $\lceil t_0 k \rceil$ *values of* $i \in [0,k]$.

**Proof.** By Equation (4.5), we get

$$\|y_k\| \le M\|\delta_k\| \le M^2 \|s_k\|.$$

Using this and Equation (4.6), we obtain

$$\frac{\|y_k\|^2}{s_k^T y_k} \le \frac{M^4 \|s_k\|^2}{m_1 \|s_k\|^2} \le \frac{M^4}{m_1} = M_1, M_1 = \frac{M^4}{m_1}. \quad (4.13)$$

From Equation (3.2), we have

$$\begin{aligned} Tr(B_{k+1}) &= Tr(B_k) - (1-\phi_k)\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} \\ &+ \left(1+\phi_k \frac{s_k^T B_k s_k}{s_k^T y_k}\right)\frac{\|y_k\|^2}{s_k^T y_k} - 2\phi_k \frac{s_k^T B_k y_k}{s_k^T y_k}, \end{aligned} \quad (4.14)$$

where $Tr(B_k)$ denote the trace of $B_k$ By Lemma 4.2, we know there exists a positive integer $k'$, when $k \ge k'$, Equation (4.6) holds. Let us now define $N_k$ by

$$N_k = \{k \mid k \ge k' holds\}$$

Denote

$$I_1 = \{k \mid 0 \le \phi_k \le 1-\ell, k \in N_k\},$$

$$I_2 = N_k - I_1 = \{k \mid (1-v)\phi_k^c \le \phi_k < 0, k \in N_k\}.$$

Consider the following two cases.

1) $k \in I_1$. Equation (4.13) indicates that

$$\frac{\|y_k\|^2}{s_k^T y_k} \cdot \frac{s_k^T B_k s_k}{s_k^T y_k} \Bigg/ \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} \le M_1 \frac{\left(s_k^T B_k s_k\right)^2}{s_k^T y_k \|B_k s_k\|^2} \quad (4.15)$$

and

$$\frac{\left|s_k^T B_k y_k\right|}{s_k^T y_k} \Bigg/ \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} \le \frac{\|y_k\| s_k^T B_k s_k}{s_k^T y_k \|B_k s_k\|} \le \sqrt{M_1} \frac{s_k^T B_k s_k}{\sqrt{s_k^T y_k}\|B_k s_k\|} \quad (4.16)$$

Using $B_k$ is positive definite, Equations (4.1), (4.14)-(4.16), we have

$$Tr(B_{k+1}) \le Tr(B_k) - \frac{\ell}{2}\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + M_1 \quad (4.17)$$

holds for all $k \in I_1$.

2) $k \in I_2$. According to Equations (4.1), (4.13), and (4.14), it is easy to get

$$Tr(B_{k+1}) \le Tr(B_k) - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + M_1,$$

this means that Equation (4.17) also holds in this case. So the relation Equation (4.17) holds for these two cases. Define the Rayleigh quotient

$$p = \frac{s_k^T B_k s_k}{s_k^T s_k}, \quad (4.18)$$

and the function

$$\phi(B) = Tr(B) - \ln(\det(B)), \quad (4.19)$$

Where ln denotes the logarithm, and $B$ is any positive definite matrix.

By Equations (3.1), (4.1) and Lemma 4.4, we can deduce that

$$\begin{aligned} \det(B_{k+1}) &= \det(B_k)\frac{y_k^T s_k}{s_k^T B_k s_k}\{1+\phi_k(u_k - 1)\} \\ &\ge \det(B_k)\frac{y_k^T s_k}{s_k^T B_k s_k} v. \end{aligned} \quad (4.20)$$

From Equations (4.17), (4.19), (4.20), the definitions (4.11) and (4.18), we have

$$\begin{aligned} \varphi(B_{k+1}) &\le \varphi(B_k) - \frac{\ell}{2}\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + M_1 - \ln\left(\frac{s_k^T y_k}{s_k^T B_k s_k} v\right) \\ &= \varphi(B_k) - \frac{\ell}{2}\left(\frac{\|B_k s_k\| \cdot \|s_k\|}{s_k^T B_k s_k}\right)^2 \frac{s_k^T B_k s_k}{s_k^T s_k} \\ &+ M_1 - \ln\left(\frac{s_k^T y_k}{s_k^T s_k} \cdot \frac{s_k^T s_k}{s_k^T B_k s_k} v\right) \\ &= \varphi(B_k) + M_1 - \ln\frac{s_k^T y_k}{s_k^T s_k} v - \frac{\ell}{2} \cdot \frac{p_k}{\cos^2\theta_k} + \ln p_k \\ &= \varphi(B_k) + M_1 + \ln\frac{2}{\ell} - 1 - \ln\frac{s_k^T y_k}{s_k^T s_k} v \\ &+ \ln\cos^2\theta_k + \left(1 - \frac{\ell}{2}\frac{p_k}{\cos^2\theta_k} + \ln\frac{\ell}{2}\frac{p_k}{\cos^2\theta_k}\right). \end{aligned} \quad (4.21)$$

Combining this and Equation (4.6), we get

$$\begin{aligned} \varphi(B_{k+1}) &\le \varphi(B_k) + \left(M_1 + \ln\frac{2}{\ell} - 1 + \ln v m_1\right)\left(k - k' + 1\right) \\ &+ \sum_{i=k'}^{k}\left[\ln\cos^2\theta_i + 1 - \frac{\ell}{2} \cdot \frac{p_i}{\cos^2\theta_i} + \ln\frac{\ell}{2} \cdot \frac{p_i}{\cos^2\theta_i}\right] \end{aligned}$$

Define $\eta_i \ge 0$, by

$$\eta_i = -\ln\cos^2\theta_i - \left(1 - \frac{\ell}{2}\frac{p_i}{\cos^2\theta_i} + \ln\frac{\ell}{2}\frac{p_i}{\cos^2\theta_i}\right). \quad (4.22)$$

Since $\varphi(B_{k+1}) > 0$ [or see [46]] we have

$$\frac{1}{k-k^{'}+1}\sum_{j=k^{'}}^{k}\eta_i < \frac{\varphi(B_k)}{k-k^{'}+1} + \left(M_1 + \ln\frac{2}{\ell} - 1 - \ln vm_1\right). \quad (4.23)$$

Let us define $\eta_i$ to be a set consisting of $\lceil t_0(k-k^{'})\rceil$ indices corresponding to the $\lceil t_0(k-k^{'})\rceil$ smallest values of $\eta_i$ for $k^{'} \le i \le k$, and let $\eta_{\max}$ denote the largest of the $\eta_i$ for $i \in J_k$ Then we get

$$\frac{1}{k-k^{'}+1}\sum_{j=k^{'}}^{k}\eta_i < \frac{1}{k-k^{'}+1}\left(\eta_{\max} + \sum_{i=k^{'}, i\notin J_k}^{k}\eta_i\right).$$

$$\ge \eta_{\max}(1-t_0).$$

Therefore, by Equation (4.23), we have, for all $i \in J_k$

$$\eta_i < \frac{1}{1-t_0}\left(\varphi(B_k) + M_1 + \ln\frac{2}{\ell} - 1 - \ln vm_1\right) \equiv \varepsilon_0 \quad (4.24)$$

Since the term inside the brackets in Equation (4.22) is less than or equal to zero, we conclude from Equations (4.22) and (4.24) that for all $i \in J_k$

$$-\ln\cos^2\theta_i < \varepsilon_0$$

Thus, we get

$$\cos\theta_i > e^{-t_0/2} \equiv \beta_1 \quad (4.25)$$

According to Equations (4.22) and (4.24), for all $i \in J_k$ we have,

$$1 - \frac{\ell}{2}\cdot\frac{p_i}{\cos^2\theta_i} + \ln\frac{\ell}{2}\cdot\frac{p_i}{\cos^2\theta_i} > -\varepsilon_0.$$

Note the function

$$w(t) = 1 - t + \ln t, \quad (4.26)$$

is nonpositive for all $t > 0$, achiexes its maximum value at $t = 0$, and satisfies $w(t) \to -\infty$ both as $t \to \infty$ and $t \to 0$. Then it follows that for all $i \in J_k$

$$\beta_3 \ge \frac{p_i}{\cos^2\theta_i} \ge \beta_2^{'} > 0,$$

For some constants $\beta_2^{'}$ and $\beta_3$. By Equation (4.25), we get

$$\beta_2 \equiv \beta_1^2\beta_2^{'} \le p_i \le \beta_3$$

Using $\frac{\|B_i s_i\|}{\|s_i\|} = \frac{p_i}{\cos\theta_i}$, we obtain for all $i \in J_k$,

$$\beta_2 \le \frac{\|B_i s_i\|}{\|s_i\|} \le \frac{\beta_3}{\beta_1}.$$

Since $k^{'}$ is a fixed integer and $B_i$ are positive definite, we can take smaller $\beta_1, \beta_2$ and larger $\beta_3$ if necessary so that this lemma holds for all $i \le k^{'}$ Therefore Equation (4.12) holds for at least $\lceil t_0 k\rceil$ indices $i \in [0, k]$. The proof is complete.□

Let $N = \{i \mid (4.12)\, holds\}$.

Similar to [36], it is not difficult to get the global convergence theorem of Algorithm 1. So we only state as follows but omit the proof.

**Theorem 4.1** *Let Assumption A and Equation* (4.1) *hold. Then the sequence* $\{x_k\}$ *generated by the Gauss-Newton- based Broyden's class Algorithm. Then*

$$\lim_{k\to\infty}\inf\|g_k\| = 0. \quad (4.27)$$

## 5. Numerical Results

In this section, we report results of some numerical experiments with the proposed method. We will test two practically statistical problems to show the efficiency of Algorithm 1.

**Problem 1.** In **Table 1**, there is data of the age $x$ and the average height $H$ of a pine tree:

Our objective is to find out the approximate function between the demand and the price, namely, we need to find the regression equation of $x$ to the $h$ .It is easy to see that the age $x$ and the average height $H$ are parabola relations. Denote the regression function by $h = \beta_0 + \beta_1 x + \beta_2 x^2$ where $\beta_0$, $\beta_1$, and $\beta_2$ are the regression parameters. Using least squares method, we need to solve the following problem

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

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

$$\min_{\beta\in\Re^3} f(\beta) = \sum_{i=1}^{n}\left[h_i - \beta\left(1, x_i, x_i^2\right)^T\right]^2. \quad (5.1)$$

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.

In the experiment, all codes were written in **MATLAB 7.5** and run on **PC** with **2.60 GHz CPU** processor and **480 MB** memory and **Windows XP** operation system. In the experiments, the parameters in Algorithm 1 were chosen as $r = 0.1$, $\rho = \sqrt{0.85}$, $\sigma_1 = \sigma_2 = 10^{-4}$, $\alpha_{-1}$ = 0.0001 and $\varepsilon_k = k^{-2}$, where $k$ is the number of iteration. The initial matrix $B_0$ was always set to be the unit matrix. We will stop the program if the condition $\|g(\beta)\| \le le-5$ is satisfied.

In order to show the efficiency of these algorithms, the residuals of sum of squares is defined by

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

Where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_n X_{in}, i = 1, 2, \cdots, n,$ and $\hat{y}_i = \hat{\beta}_0, \hat{\beta}_1 \cdots, \hat{\beta}_n$ are the parameters when the program is stopped or the solution is obtained from one way. Let

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

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 [48]. In **Table 2**, $DFP$ stands for the Formula (3.2) in Algorithm 1 where $\phi_k = 1$, and $BFGS$ stands for the Formula (3.2) in Algorithm 1 where $\phi_k = 0$.

The columns of the **Table 2** have the following meaning:

$\beta^*$: the approximate solution from the method of extreme value of calculus or some software.

$\overset{\backslash}{\beta}$: the solution as the program is terminated. $\check{\beta}$: the initial point. NI: the total number of iterations. $\varepsilon_*$: the relative error between $RMS_p\left(\beta^*\right)$ and $RMS_p\left(\beta^{\backslash}\right)$ defined by

**Table 1. The data of 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 2. Test result for problem 1.**

| $\beta^*$ | $\check{\beta}$ | $\overset{\backslash}{\beta}$ | $RMS_p\left(\overset{\backslash}{\beta}\right)$ | $RMS_p\left(\beta^*\right)$ | $\varepsilon_*$ | NI |
|---|---|---|---|---|---|---|
| DFP | (-1,30,-5) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 11 |
| | (1000,1000,1000) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 6 |
| | (0,0,0) | (-1.331363,3.461743,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 10 |
| | (-10,100,-1000) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 11 |
| | (-10,-100,-1000) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 9 |
| | (10,-100,1000) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 12 |
| | (500,-600,700) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 6 |
| | (1,2,3) | (-1.331363,3.461743,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 11 |
| | (-1,-2,-3) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 6 |
| | (3,2,1) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 10 |
| BFGS | (-1,30,-5) | (-1.331363,3.461739,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 10 |
| | (1000,1000,1000) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 8 |
| | (0,0,0) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 10 |
| | (-10,100,-1000) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 8 |
| | (-10,-100,-1000) | (-1.331363,3.461747,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 9 |
| | (10,-100,1000) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 8 |
| | (500,-600,700) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 9 |
| | (1,2,3) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 10 |
| | (-1,-2,-3) | (-1.331363,3.461742,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 9 |
| | (3,2,1) | (-1.331363,3.461743,-0.108712) | 0.171712 | 0.183900 | 6.627449% | 10 |

$$\varepsilon_* = \frac{RMS_p\left(\beta^*\right) - RMS_P\left(\overset{\backslash}{\beta}\right)}{RMS_p\left(\beta^*\right)}.$$

For Problem 1, the above problems (5.2) can be solved by extreme value of calculus. Then we get $\beta^* = \left(-1.33, 3.46, -0.11\right)$ in **Table 2**. Here we also solve these two problems by Algorithm 1. These numerical results of **Table 2** indicate that Algorithm 1 is 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. Moreover, the initial points don not influence that the sequence $\{x_k\}$ converges to one solution $x_*$ our proposed method.

## 6. References


[1] D. M. Bates and D. G. Watts, "Nonlinear Regression Analysis and Its Applications," John Wiley & Sons, New York, 1988. doi:10.1002/9780470316757

[2] S. Chatterjee and M. Machler, "Robust Regression: A Weighted Least Squares Approach, Communications in Statistics," *Theorey and Methods*, Vol. 26, No. 6, 1997, pp. 1381-1394. doi:10.1080/03610929708831988

[3] R. Christensen, "Analysis of Variance, Design and Regression: Applied Statistical Methods," Chapman and Hall, New York, 1996.

[4] N. R. Draper and H. Smith, "Applied Regression Analysis," 3rd Edtion, John Wiley & Sons, New York, 1998.

[5] F. A. Graybill and H. K. Iyer, "Regression Analysis: Concepts and Applications," Duxbury Press, Belmont, 1994.

[6] R. F. Gunst and R. L. Mason, "Regression Analysis and Its Application: A Data-Oriented Approach," Marcel Dekker, New York, 1980.

[7] R. H. Myers, "Classical and Modern Regression with Applications," 2nd Edtion, PWS-Kent Publishing Company, Boston, 1990.

[8] R. C. Rao, "Linear Statistical Inference and Its Applications," John Wiley & Sons, New York, 1973. doi:10.1002/9780470316436

[9] D. A. Ratkowsky, "Nonlinear Regression Modeling: A Unified Practical Approach," Marcel Dekker, New York, 1983.

[10] D. A. Ratkowsky, "Handbook of Nonlinear Regression Modeling," Marcel Dekker, New York, 1990.

[11] A. C. Rencher, "Methods of Multivariate Analysis," John Wiley & Sons, New York, 1995.

[12] G. A. F. Seber and C. J. Wild, "Nonlinear Regression," John Wiley & Sons, New York, 1989. doi:10.1002/0471725315

[13] A. Sen and M. Srivastava, "Regression Analysis: Theory, Methods, and Applications," Springer-Verlag, New York, 1990.

[14] J. Fox, "Linear Statistical Models and Related Methods," John Wiley & Sons, New York, 1984.

[15] S. Haberman and A. E. Renshaw, "Generalized Linear Models and Actuarial Science," The Statistician, Vol. 45, No. 4, 1996, pp. 407-436. doi:10.2307/2988543

[16] S. Haberman and A. E. Renshaw, "Generalized Linear Models and Excess Mortality from Peptic Ulcers," *Insurance: Mathematics and Economics*, Vol. 9, No. 1, 1990, pp. 147-154. doi:10.1016/0167-6687(90)90012-3

[17] R. R. Hocking, "The Analysis and Selection of Variables in Linear Regression," *Biometrics*, Vol. 32, No. 1, 1976, pp. 1-49. doi:10.2307/2529336

[18] P. McCullagh and J. A. Nelder, "Generalized Linear Models," Chapman and Hall, London, 1989.

[19] J. A. Nelder and R. J. Verral, "Credibility Theory and Generalized Linear Models." *ASTIN Bulletin*, Vol. 27, No. 1, 1997, pp. 71-82. doi:10.2143/AST.27.1.563206

[20] M. Raydan, "The Barzilai and Borwein Gradient Method for the Large Scale Unconstrained Minimization Problem," *SIAM Journal on Optimization*, Vol. 7, No. 1, 1997, pp. 26-33. doi:10.1137/S1052623494266365

[21] J. Schropp, "A Note on Minimization Problems and Multistep Methods," *Numerical Mathematics*, Vol. 78, 1997, pp. 87-101. doi:10.1007/s002110050305

[22] J. Schropp, "One-Step and Multistep Procedures for Constrained Minimization Problems," *IMA Journal of Numerical Analysis*, Vol. 20, No. 1, 2000, pp. 135-152. doi:10.1093/imanum/20.1.135

[23] D. J. Van Wyk, "Differential Optimization Techniques," *Applied Mathematical Modelling*, Vol. 8, 1984, pp. 419-424. doi:10.1016/0307-904X(84)90048-9

[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*, Vol. 114, No. 2, 2000, pp. 367-386. doi:10.1016/S0377-0427(99)00276-9

[25] G. Yuan and X. Lu, "A New Line Search Method with Trust Region for Unconstrained Optimization," *Communications on Applied Nonlinear Analysis*, Vol. 15, 2008, No. 2, pp. 35-49.

[26] G. Yuan and Z. Wei, "New Line Search Methods for Unconstrained Optimization," *Journal of the Korean Statistical Society*, Vol. 38, No. 1, 2009, pp. 29-39. doi:10.1016/j.jkss.2008.05.004

[27] G. Yuan and X. Lu, "A Modified PRP Conjugate Gradient Method," *Annals of Operations Research*, Vol. 166, No. 1, 2009, pp. 73-90. doi:10.1007/s10479-008-0420-4

[28] 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. doi:10.1007/s11590-008-0086-5

[29] G. Yuan, X. Lu and Z. Wei, "A Conjugate Gradient Me-


 

thod with Descent Direction for Unconstrained Optimization," *Journal of Computational and Applied Mathematics*, Vol. 233, No. 2, 2009, pp. 519-530. doi:10.1016/j.cam.2009.08.001

[30] G. Yuan, "A Conjugate Gradient Method for Unconstrained Optimization Problems," *International Journal of Mathematics and Mathematical Sciences*, Vol. 2009, 2009, pp. 1-14. doi:10.1155/2009/329623

[31] G. Yuan and Z. Wei, "A Nonmonotone Line Search Method for Regression Analysis," *Journal of Service Science and Management*, Vol. 2, No. 1, 2009, pp. 36-42. doi:10.4236/jssm.2009.21005

[32] A. Griewank, "The 'Global' Convergence of Broyden-Like Methods with a Suitable Line Search," *Journal of the Australian Mathematical Society. Series B*, Vol. 28, No. 1, 1986, pp. 75-92. doi:10.1017/S0334270000005208

[33] D. T. Zhu, "Nonmonotone Backtracking Inexact Quasi-Newton Algorithms for Solving Smooth Nonlinear Equations," *Applied Mathematics and Computation*, Vol. 161, No. 3, 2005, pp. 875-895. doi:10.1016/j.amc.2003.12.074

[34] J. Y. Fan, "A Modified Levenberg-Marquardt Algorithm for Singular System of Nonlinear Equations," *Journal of Computational Mathematics*, Vol. 21, No. 5, 2003, pp. 625-636.

[35] Y. Yuan, "Trust Region Algorithm for Nonlinear Equations," *Information*, Vol. 1, 1998, pp. 7-21.

[36] D. Li and M. Fukushima, "A Global and Superlinear Convergent Gauss-Newton-Based BFGS Method for Symmetric Nonlinear Equations," *SIAM Journal on Numerical Analysis*, Vol. 37, No. 1, 1999, pp. 152-172. doi:10.1137/S0036142998335704

[37] Z. Wei, G. Yuan and Z. Lian, "An Approximate Gauss-Newton-Based BFGS Method for Solving Symmetric Nonlinear Equations," *Guangxi Sciences*, Vol. 11, No. 2, 2004, pp. 91-99.

[38] G. Yuan and X. Li, "An Approximate Gauss-Newton-Based BFGS Method with Descent Directions for Solving Symmetric Nonlinear Equations," *OR Transactions*, Vol. 8, No. 4, 2004, pp. 10-26.

[39] G. Yuan and X. Lu, "A Nonmonotone Gauss-Newton-Based BFGS Method for Solving Symmetric Nonlinear Equations," *Journal of Lanzhou University*, Vol. 41, 2005, pp. 851-855.

[40] G. Yuan abd X. Lu, "A New Backtracking Inexact BFGS Method for Symmetric Nonlinear Equations," *Computer and Mathematics with Application*," Vol. 55, No. 1, 2008, pp. 116-129. doi:10.1016/j.camwa.2006.12.081

[41] G. Yuan, X. Lu and Z. Wei, "BFGS Trust-Region Method for Symmetric Nonlinear Equations," *Journal of Computational and Applied Mathematics*, Vol. 230, No. 1, 2009, pp. 44-58. doi:10.1016/j.cam.2008.10.062

[42] G. Yuan, Z. Wang and Z. Wei, "A Rank-One Fitting Method with Descent Direction for Solving Symmetric Nonlinear Equations," *International Journal of Communications, Network and System Sciences*, Vol. 2, No. 6, 2009, pp. 555-561. doi:10.4236/ijcns.2009.26061

[43] G. Yuan, S. Meng and Z. Wei, "A Trust-Region-Based BFGS Method with Line Search Technique for Symmetric Nonlinear Equations," *Advances in Operations Research*, Vol. 2009, 2009, pp. 1-20. doi:10.1155/2009/909753

[44] G. Yuan and X. Li, "A Rank-One Fitting Method for Solving Symmetric Nonlinear Equations," *Journal of Applied Functional Analysis*, Vol. 5, No. 4, 2010, pp. 389-407.

[45] G. Yuan, Z. Wei and X. Lu, "A Nonmonotone Trust Region Method for Solving Symmetric Nonlinear Equations," *Chinese Quarterly Journal of Mathematics*, Vol. 24, No. 4, 2009, pp. 574-584.

[46] R. Byrd and J. Nocedal, "A Tool for the Analysis of Quasi-Newton Methods with Application to Unconstrained Minimization," *SIAM Journal on Numerical Analysis*, Vol. 26, No. 3, 1989, pp. 727-739. doi:10.1137/0726042

[47] D. Xu, "Global Convergence of the Broyden's Class of Quasi-Newton Methods with Nonomonotone Linesearch," *ACTA Mathematicae Applicatae Sinica, English Series*, Vol. 19, No. 1, 2003, pp.19-24. doi:10.1007/s10255-003-0076-4

[48] S. Chatterjee, A. S. Hadi and B. Price, "Regression Analysis by Example," 3rd Edition, John Wiley & Sons, New York, 2000.

 