A MODIFIED REGULARIZED NEWTON METHOD FOR UNCONSTRAINED NONCONVEX OPTIMIZATION

Heng Wang*, Mei Qin & Haibo Wang
College of Science, University of Shanghai for Science and Technology, Shanghai, China, 200093
*Email: wanghengusst@126.com

ABSTRACT

In this paper, we present a modified regularized Newton method for the unconstrained nonconvex optimization by using trust region technique. We show that if the gradient and Hessian of the objective function are Lipschitz continuous, then the modified regularized Newton method (M-RNM) has a global convergence property. Numerical results show that the algorithm is very efficient.

Keywords: Regularized Newton method; Trust region method; Unconstrained nonconvex optimization; Global convergence

1. INTRODUCTION

We consider the unconstrained optimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, whose gradient $\nabla f$ and Hessian $\nabla^2 f$ are denoted by $g(x)$ and $H(x)$ respectively. Throughout this paper, we assume that the solution set of (1.1) is nonempty and denoted by $S^*$, and in all cases $\| \|$ refers to the 2-norm.

It is well known that $f(x)$ is convex if and only if $H(x)$ is symmetric positive semidefinite for all $x \in \mathbb{R}^n$. Moreover, if $f(x)$ is convex, then $x \in S^*$ if and only if $x$ is a solution of the system of nonlinear equations

$$g(x) = 0.$$  

Hence, we could get the minimizer of $f(x)$ by solving (1.2) [1]-[3]. The Newton method is one of a efficient solution method. At every iteration, it computes the trial step

$$d_k^N = -H_k^{-1}g_k,$$

where $g_k = g(x_k)$ and $H_k = H(x_k)$. As we know, if $H_k$ is Lipschitz continuous and nonsingular at the solution, then the Newton method has quadratic convergence. However, this method has an obvious disadvantage when the $H_k$ is singular or near singular.

To overcome the difficulty caused by the possible singularity of $H_k$, [4] proposed a regularized Newton method, where the trial step is the solution of the linear equations

$$(H_k + \mu_k I)d = -g_k,$$

where $I$ is the identity matrix. $\mu_k$ is a positive parameter which is updated from iteration to iteration. The parameter $\mu_k$ is introduced to overcome the difficulties caused by the singularity or near singularity of the Hessian $H_k$.

Now we need to consider another question, ”how to choose the modified regularized parameter $\mu_k$?” which will play important roles not only in theoretical analysis but also in numerical experiments. Yamashita and Fukushima [5] chose $\lambda_k = \| g_k \|^2$ and showed that the regularized Newton method has quadratic convergence under the local error bound.
condition which is weaker than nonsingularity. Fan and Yuan [6] took \( \lambda_k = \|g_k\|^\delta \) with \( \delta \in [1,2] \) and proved that the Levenberg-Marquardt method preserves the quadratic convergence under the same conditions. In this paper we will choose the parameter \( \mu_k \) as \( \mu_k = \alpha_i \Lambda_k + \lambda_k \|g_k\| \), where \( \alpha_i \geq 1 \) is a positive constant, \( \Lambda_k = \max(0, -\lambda_{\min}(H_k)) \). \( \lambda_k > 0 \) is updated from iteration to iteration by a trust region technique.

In most past studies [7]-[9] for the regularized Newton method, the convergence properties have been discussed only when \( f \) is convex. In this paper, we propose a modified Newton method for (1.1) whose objective function \( \mu_k = \alpha_i \Lambda_k + \lambda_k \|g_k\| \), where \( \alpha_i \geq 1 \) is a positive constant, \( \lambda_k > 0 \) is updated from iteration to iteration by a trust region technique.

We extend the regularized Newton method (1.4) to the unconstrained nonconvex optimization. At the \( k \)-th iteration of the M-RNM, we set regularized parameter \( \mu_k \) as \( \mu_k = \alpha_i \Lambda_k + \lambda_k \|g_k\| \), where \( \alpha_i \geq 1 \) and \( \lambda_k > 0 \). From the definition of \( \Lambda_k \), the matrix \( H_k + \alpha_i \Lambda_k I \) is positive semidefinite even if \( f \) is nonconvex. Therefore, if \( \|g_k\| \neq 0 \), then \( H_k + \mu_k I = H_k + \alpha_i \Lambda_k I + \lambda_k \|g_k\|I \geq 0 \), we can use regularized Newton method to solve the problem of (1.1).

The main scheme of the modified regularized Newton method for unconstrained nonconvex optimization is given as follows. At every iteration, it solves the linear equations

\[
(H_k + \mu_k I) d = -g_k
\]

(1.7)

to obtain the Newton step \( d_k \), where \( \mu_k = \alpha_i \Lambda_k + \lambda_k \|g_k\| \), and then solves the linear equations

\[
(H_k + \mu_k I) d = -g(y_k) \quad \text{with} \quad y_k = x_k + d_k
\]

(1.8)

to obtain the approximate Newton step \( \tilde{d}_k \).

The purpose of this paper is to investigate whether the proposed method for unconstrained nonconvex optimization has global convergence.

The paper is organized as follows. In section 2, we present a new modified regularized Newton algorithm by using trust region technique, then prove the global convergence. In section 3, numerical results are given. Finally, we conclude the paper in the section 4.

2. THE ALGORITHM AND GLOBAL CONVERGENCE

In this section, we first present the new modified regularized Newton algorithm by using trust region technique, then prove the global convergence. First, we give the modified regularized Newton algorithm.

Let \( d_k \) and \( \tilde{d}_k \) be given by (1.7) and (1.8), respectively. Since the matrix \( d_k + \tilde{d}_k \) is symmetric and positive definite, \( d_k \) is a descent direction of \( f(x) \) at \( x_k \), but \( d_k + \tilde{d}_k \) may not be. Hence we prefer to use a trust region technique.
Define the actual reduction of $f(x)$ at the $k$-th iteration as

$$A_{red_k} = f(x_k) - f(x_k + d_k + \tilde{d}_k).$$

(2.1)

Note that the Newton step $d_k$ is the minimizer of the problem:

$$\min_{d \in \mathbb{R}^n} \phi_{k,1}(d) = \frac{1}{2} d^T H_k d + g_k^T d + \frac{1}{2} \mu_k \|d\|^2.$$

If we let

$$\Delta_{k,1} = \|d_k\| = \left\| - (H_k + \mu_k I)^{-1} g_k \right\|,$$

then it can be proved [1] that $d_k$ is also a solution of the trust region problem:

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} d^T H_k d + g_k^T d, \text{ s.t. } \|d\| \leq \Delta_{k,1}.$$

By the famous result given by Powell in [11], we know that

$$\phi_{k,1}(0) - \phi_{k,1}(d_k) \geq \frac{1}{2} \|g_k\| \min \left\{ \|d_k\|, \frac{\|g_k\|}{\|H_k\|} \right\}.$$

(2.2)

Similar to $d_k$, $\tilde{d}_k$ is not only the minimizer of the problem:

$$\min_{d \in \mathbb{R}^n} \phi_{k,1}(d) = \frac{1}{2} d^T H_k d + g_k^T d + \frac{1}{2} \mu_k \|d\|^2,$$

but also the solution of the following trust region problem:

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} d^T H_k d + g_k^T d, \text{ s.t. } \|d\| \leq \Delta_{k,2},$$

where

$$\Delta_{k,2} = \|\tilde{d}_k\| = \left\| - (H_k + \mu_k I)^{-1} g_k \right\|.$$

Therefore we also have

$$\phi_{k,2}(0) - \phi_{k,2}(d_k) \geq \frac{1}{2} \|g_k\| \min \left\{ \|\tilde{d}_k\|, \frac{\|g_k\|}{\|H_k\|} \right\}.$$

(2.3)

Based on the inequalities (2.2) and (2.3), it is reasonable for us to define the new predicted reduction as

$$\text{Pred}_k = \phi_{k,1}(0) - \phi_{k,1}(d_k) + \phi_{k,2}(0) - \phi_{k,2}(\tilde{d}_k),$$

(2.4)

which satisfies

$$\text{Pred}_k \geq \frac{1}{2} \|g_k\| \min \left\{ \|d_k\|, \frac{\|g_k\|}{\|H_k\|} \right\} + \frac{1}{2} \|g_k\| \min \left\{ \|\tilde{d}_k\|, \frac{\|g_k\|}{\|H_k\|} \right\}.$$

(2.5)

The ratio of the actual reduction to the predicted reduction

$$r_k = \frac{A_{red_k}}{\text{Pred}_k},$$

(2.6)

plays a key role in deciding whether to accept the trial step and how to adjust the regularized parameter.
The regularized Newton algorithm with correction for unconstrained nonconvex optimization problems is stated as follows.

(0,0)→(13,0);

**Algorithm 2.1**

Step 1. Given \( x_0 \in \mathbb{R}^n, \ \varepsilon \geq 0, \ \lambda_0 > m > 0, \ 0 < c_0 \leq c_1 \leq c_2 < 1, \ 0 < p_1 < 1 < p_2, \ \alpha_i \geq 1, \ k := 0. \)

Step 2. If \( \| g_k \| \leq \varepsilon \), then stop. Otherwise go to step 3.

Step 3. Compute \( \Lambda_k = \max(0,-\lambda_{\text{min}}(H_k)), \ v_k = \lambda_k \| g_k \|. \)

Solve

\[
(H_k + \alpha_i \Lambda_k I + v_k I) d = -g_k. 
\]

(2.7)

Solve

\[
(H_k + \alpha_i \Lambda_k I + v_k I) d = -g(y_k),
\]

(2.8)

to obtain \( d_k \).

Set

\[ y_k = x_k + d_k. \]

Set

\[ s_k = d_k + \square. \]

Step 4. Compute \( r_k = \frac{A_{\text{red}}}{P_{\text{red}}}. \)

Set

\[ x_{k+1} = \begin{cases} 
   x_k + s_k, & \text{if } r_k \geq c_0, \\
   x_k, & \text{otherwise.}
\end{cases} \]

(2.9)

Step 5. Update \( \lambda_{k+1} \) as

\[
\lambda_{k+1} = \begin{cases} 
   p_2 \lambda_k, & \text{if } r_k < c_1, \\
   \lambda_k, & \text{if } r_k \in [c_1, c_2], \\
   \max \{ p_1 \lambda_k, m \}, & \text{if } r_k > c_2.
\end{cases}
\]

(2.10)

Set \( k := k + 1 \) and go step 2.

In Algorithm 2.1, the given positive constant \( m \) is the lowerbound of \( \lambda_k \). It plays the role in preventing the step from being too large when the sequence is near the minimizer set.

Before discussing the global convergence of the algorithm above, we make the following assumption.

**Assumption 2.2** \( g(x) \) and \( H(x) \) are both Lipschitz continuous, that is, there exists a constant \( L_1 > 0, \ L_2 > 0 \) such that
\[ \| g(y) - g(x) \| \leq L_2 \| y - x \|, \forall x, y \in \mathbb{R}^n \]  
\[(2.11)\]

and
\[ \| H(y) - H(x) \| \leq L_4 \| y - x \|, \forall x, y \in \mathbb{R}^n. \]  
\[(2.12)\]

It follows from (2.12) that
\[ \| g(y) - g(x) - H(x)(y - x) \| \leq L_4 \| y - x \|^2, \forall x, y \in \mathbb{R}^n. \]  
\[(2.13)\]

The following lemma given below shows the relationship between the positive semidefinite matrix and symmetric positive semidefinite matrix.

**Lemma 2.3** A real-valued matrix is positive semidefinite if and only if \( (A + A^T)/2 \) is positive semidefinite.

**Proof.** See [1]

Next, we give the bounds of a positive definite matrix and its inverse.

**Lemma 2.4** Suppose \( A \) is symmetric positive definite. Then,
\[ \| A + \varphi I \| \geq \varphi \]
and
\[ \| (A + \varphi I)^{-1} \| \leq \varphi^{-1} \]
hold for any \( \varphi > 0 \).

**Proof.** It follows from Lemma 2.3 and the definition of the 2-norm that
\[ \| A + \varphi I \| = \sqrt{\lambda_{\text{max}}((A + \varphi I)^T(A + \varphi I))} \]
\[ = \sqrt{\lambda_{\text{max}}(A^TA + \varphi(A + A^T) + \varphi^2 I)} \]
\[ \geq \sqrt{\lambda_{\text{max}}(\varphi^2 I)} \]
\[ = \varphi, \]
where \( \lambda_{\text{max}}((A + \varphi I)^T(A + \varphi I)) \) means the largest eigenvalue of \( (A + \varphi I)^T(A + \varphi I) \). Similarly, we have
\[ \| (A + \varphi I)^{-1} \| = \sqrt{\lambda_{\text{max}}((A + \varphi I)^{-1}(A + \varphi I)^{-1})} \]
\[ = \sqrt{\lambda_{\text{max}}((A + \varphi I)^{-1}(A + \varphi I)^{-1})} \]
\[ = \sqrt{\lambda_{\text{min}}(A^TA + \varphi(A + A^T) + \varphi^2 I)} \]
\[ \leq \varphi^{-1}. \]

This completes the proof.

So we have the following results.

**Theorem 2.5** Under the conditions of Assumption 2.2, if \( f \) is bounded below, then Algorithm 2.1 terminates in finite iterations or satisfies
\[ \lim \inf_{k \to \infty} \| g_k \| = 0. \]  
\[(2.14)\]

**Proof.** We prove by contradiction. If the theorem is not true, then there exists a positive \( \tau \) and an integer \( \tilde{k} \) such that
\[ \| g_{\tilde{k}} \| \geq \tau, \forall k \geq \tilde{k}. \]  
\[(2.15)\]
Without loss of generality, we can suppose \( k = 1 \). Set \( T = \{ k \mid x_k \neq x_{k+1}\} \). Then 
\[ \{1,2,\ldots\} = T \cup \{ k \mid x_k = x_{k+1}\}. \]

Now we will analysis in two cases whether \( T \) is finite or not.

Case (1): \( T \) is finite. Then there exists an integer \( k_1 \) such that 
\[ x_{k_1} = x_{k_1+1} = x_{k_1+2} = \cdots. \]

By (2.9), we have 
\[ r_k < c_0, \forall k \geq k_1. \]

Therefore by (2.10) and (2.15), we deduce 
\[ \lambda_k \to \infty, v_k \to \infty. \quad (2.16) \]

Since \( x_{k+1} = x_k, \forall k \geq k_1 \), we get from (2.7) and (2.16) that 
\[ \| d_k \| = \left\| (H_k + \alpha I_k I + \lambda_k \|g\| I)^{-1} g_k \right\| \to 0. \quad (2.17) \]

From (2.8), we obtain 
\[
\left\| \hat{d}_k \right\| = \left\| - \left( H_k + \alpha I_k I + \lambda_k \|g\| I \right)^{-1} g_k \right\| \\
\leq \left\| \left( H_k + \alpha I_k I + \lambda_k \|g\| I \right)^{-1} (g(y_k) - g_k - H_k d_k) \right\| \\
+ \left\| \left( H_k + \alpha I_k I + \lambda_k \|g\| I \right)^{-1} g_k \right\| + \left( H_k + \alpha I_k I + \lambda_k \|g\| I \right)^{-1} H_k d_k \\
\leq L \lambda_k^{-1} \frac{1}{\|g\|} \left\| d_k \right\|^2 + 2\| d_k \| \\
\leq \gamma_1 \| d_k \| ,
\]

where \( \gamma_1 \) is a positive constant.

It follows from (2.1) and (2.4) that 
\[
| \text{Ared}_k - \text{Pred}_k | = \left| f(x_k) - f(x_k + d_k + \hat{d}_k) - (\phi_{k,1}(0) - \phi_{k,1}(d_k) + \phi_{k,2}(0) - \phi_{k,2}(\hat{d}_k)) \right| \\
\leq f(y_k + \hat{d}_k) - f(y_k) - \frac{1}{2} \hat{d}_k^T H_k \hat{d}_k - g(y_k)^T \hat{d}_k \\
+ f(y_k) - f(x_k) - \frac{1}{2} d_k^T H_k d_k - g_k^T d_k \\
\leq a\| d_k \|^2 + a \left( \| \hat{d}_k \|^2 \right). \quad (2.19)
\]

Moreover, from (2.5),(2.15),(2.11) and (2.17), we have 
\[
\text{Pred}_k \geq \frac{1}{2} \tau \min \left\{ \| d_k \|, \frac{\tau}{L_2} \right\} \geq \frac{1}{2} \tau \| d_k \| . \quad (2.20)
\]

for sufficiently large \( k \).

Duo to (2.19) and (2.20), we get
\[
|r_k - 1| = \left| \frac{A_{red} - Pr ed_k}{Pr ed_k} \right|
\leq \left| f(x_k) - f(x_k^* + d_k^* + d_k^*) - (\phi_{k,1}(0) - \phi_{k,1}(d_k^*) + \phi_{k,2}(0) - \phi_{k,2}(d_k^*)) \right|
+ \frac{1}{2} \tau \min \left\{ \|d_k\| \frac{\tau}{L_2} \right\}
\leq o\left(\|d_k\|^2\right) + o\left(\|d_k\|^2\right)
\|d_k\| \rightarrow 0,
\]

which implies that \( r_k \rightarrow 1 \). Hence, there exists positive constant \( \gamma_2 \) such that \( \lambda_k \leq \gamma_2 \), holds for all large \( k \), which contradicts to (2.16).

Case (2): \( T \) is infinite. Then we have from (2.5) and (2.15) that
\[
\infty > f(x_t) - \lim_{k \rightarrow \infty} f(x_k) \geq \sum_{k \in T} (f(x_k) - f(x_{k+1}))
= \sum_{k \in T} (f(x_k) - f(x_{k+1})) \geq \sum_{k \in T} (f(x_k) - f(x_{k+1})) \geq \sum_{k \in T} (f(x_k) - f(x_{k+1})) \geq \sum_{k \in T} \frac{1}{2} \|g\| \min \left\{ \|d_k\| g(H_k) \right\}
\leq \sum_{k \in T} \frac{\tau}{2} \min \left\{ \|d_k\| \frac{\tau}{L_2} \right\},
\]

which implies that
\[
\lim_{k \rightarrow \infty, k \in T} d_k = 0.
\]
The above equality together with the updating rule of (2.10) means
\[
V_k \rightarrow \infty.
\]

Similar to (2.18), it follows from (2.23) and (2.24) that
\[
\|x_k\| = \left\| d_k + d_k^* \right\| \leq (1 + \gamma_3) \|d_k\|, \forall k \in T
\]
for some positive constant \( \gamma_3 \). Then we have
\[
\|x_k\| = \left\| d_k + d_k^* \right\| \leq (1 + \gamma_3) \|d_k\|, \forall k \in T.
\]
This equality together with (2.22) yields
\[
\sum_{k \in T} \|x_k\| < \infty,
\]
which implies that
\[
x_k \rightarrow x^*.
\]
It follows from (2.7), (2.25), (2.24) and (2.18) that
\[ d_k \to 0, \quad d_k \to 0. \]  
(2.26)

Since \( (H_k + \alpha_k A_k I + \lambda_k g_k \|I\|) d = -g_k \) from (2.7), we have from (2.15), (2.11) and (2.26) that
\[ 1 \leq \frac{\|H_k\|}{\|g_k\|} \|d_k\| + \frac{\alpha_k \|A_k\|}{\|g_k\|} \|d_k\| + \lambda_k \|g_k\| \leq \frac{L_2}{\tau} \|d_k\| + \frac{\alpha_k \|A_k\|}{\tau} \|d_k\| + \lambda_k \|d_k\|, \]
which means
\[ \lambda_k \to \infty. \]  
(2.27)

By the same analysis as (2.21) we know that
\[ r_k \to 1, \]  
(2.28)

Hence, there exists a positive constant \( \gamma_k > m \) such that \( \lambda_k \leq \gamma_k \) holds for all sufficiently large \( k \), which gives a contradiction to (2.27). The proof is completed.

3. NUMERICAL EXPERIMENTS

In this section, we test the performance of the modified regularized Newton method and the regularized Newton method without correction. The function to be minimized is
\[ f(x) = \frac{1}{2} \sum_{i=1}^{n} (x_i - x_{i+1})^2 + \frac{1}{12} \sum_{i=1}^{n} \alpha_i (x_i - x_{i+1})^4 - \frac{1}{30} \sum_{i=1}^{n} \beta_i (x_i - x_{i+1})^6, \]  
(3.1)

where \( \alpha_i, \beta_i \geq 0 \) \( (i = 1, \ldots, n - 1) \) are constants. It is clear that function \( f(x) \) is nonconvex and the minimizer set of \( f(x) \) is
\[ S = \{ x \in \mathbb{R}^n \mid x_1 = x_2 = \cdots = x_n \}. \]

It can also be proved that \( \|g(x)\| \) provides a local error bound near the minimizer of \( f(x) \). The Hessian \( \nabla^2 f(x) \) is given by
\[
\nabla^2 f(x) = \begin{pmatrix}
1 & -1 \\
-1 & 2 & -1 \\
& & & \ddots \\
& & & -1 & 2 & -1 \\
& & & & -1 & 1
\end{pmatrix} + \begin{pmatrix}
b_1 - d_1 & -b_1 + d_1 \\
-b_1 + d_1 & b_1 + b_2 - d_1 - d_2 & -b_2 + d_2 \\
& & \ddots \\
& & -b_{n-2} + d_{n-2} & b_{n-2} + b_{n-1} - d_{n-2} - d_{n-1} & -b_{n-1} + d_{n-1} \\
& & & -b_{n-2} + d_{n-2} & b_{n-1} - d_{n-1}
\end{pmatrix}
\]
where \( a_i = a_i(x) = \frac{1}{3} \alpha_i (x_i - x_{i+1})^3, \) \( b_i = b_i(x) = \frac{1}{5} \beta_i (x_i - x_{i+1})^5, \) \( c_i = c_i(x) = \alpha_i (x_i - x_{i+1})^2, \)
\( d_i = d_i(x) = \beta_i (x_i - x_{i+1})^4, \) \( (i = 1, 2, \ldots, n - 1). \)
Matrix $\nabla^2 f(x)$ is indefinite for all $x$, but singular as the sum of every column is zero. Since the Hessian $H_k$ is always singular, the Newton method cannot be used to solve nonlinear equations (1.2). But by using the regularization technique, both regularized Newton method and Algorithm 2.1 work quite well.

We tested the modified regularized Newton method for problem (1.1) where $f$ is given by (3.1) with different values of $\alpha$, $\beta$ and $n$. We set $c_0 = 0.001$, $c_1 = 0.25$, $c_2 = 0.75$, $p_1 = 0.25$, $p_2 = 4$, $\alpha_1 = 1$, $\lambda_0 = 10^{-2}$ and $m = \epsilon = 10^{-5}$ for Algorithm 2.1.

We may observe that the whole sequence $\{x_k\}$ converges to $x^* = (5.5, \ldots, 5.5)^T$.

We also ran the regularized Newton algorithm (RNA) [10] without correction, that is, we do not solve the linear equations (2.8) and just set the solution of (2.7) to be the trial step $s_k$. Then, we tested the regularized Newton algorithm [10] without correction and modified regularized Newton algorithm for various of $n$, $\alpha$, $\beta$ and different choices of the starting point. The results are listed in Table 2 and Table 3. $\alpha_i$ : the selected value of $\alpha_i$; $\beta_i$ : the selected value of $\beta_i$; Dim: the dimension $n$ of the problem; $x_{0i}$: the $i$th element $x_{0i}$; md: the number of iterations required; $\|Vf\|$: the final value of $\|Vf(x_k)\|$; $x^*$: the final value of $x_k$. We use $\|Vf(x_k)\| \leq 10^{-5}$ as the stopping criterion.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$|Vf(x_k)|$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.0628</td>
</tr>
<tr>
<td>1</td>
<td>5.55e-01</td>
</tr>
<tr>
<td>2</td>
<td>6.33e-02</td>
</tr>
<tr>
<td>3</td>
<td>2.94e-04</td>
</tr>
<tr>
<td>4</td>
<td>6.63e-10</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1: Results of M-RNM on 3.1
<table>
<thead>
<tr>
<th>$\alpha_i$</th>
<th>$\beta_i$</th>
<th>$\text{Dim}$</th>
<th>$x_0$</th>
<th>$\text{md}$</th>
<th>$|\nabla f|$</th>
<th>$x^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>10</td>
<td>$i$</td>
<td>3/0</td>
<td>1.63e-07</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>2/0</td>
<td>1.85e-06</td>
<td>0.2929</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>$i$</td>
<td>13/1</td>
<td>4.28e-06</td>
<td>25.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>19/2</td>
<td>4.28e-06</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>$i$</td>
<td>16/1</td>
<td>1.62e-06</td>
<td>50.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>5/1</td>
<td>1.15e-08</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>$i$</td>
<td>49/6</td>
<td>6.31e-07</td>
<td>250.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>18/8</td>
<td>2.08e-08</td>
<td>0.0136</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>10</td>
<td>$i$</td>
<td>5/1</td>
<td>2.66e-06</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>8/2</td>
<td>5.6e-12</td>
<td>0.2929</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>$i$</td>
<td>7/3</td>
<td>3.84e-10</td>
<td>25.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>25/14</td>
<td>4.86e-09</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>$i$</td>
<td>8/2</td>
<td>4.95e-06</td>
<td>50.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>11/5</td>
<td>2.30e-06</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>$i$</td>
<td>38/19</td>
<td>3.45e-07</td>
<td>250.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>11/5</td>
<td>5.48e-06</td>
<td>0.0136</td>
</tr>
<tr>
<td>i</td>
<td>0</td>
<td>10</td>
<td>$i$</td>
<td>9/4</td>
<td>9.56e-08</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>5/1</td>
<td>9.00e-06</td>
<td>0.2929</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>$i$</td>
<td>39/16</td>
<td>2.77e-07</td>
<td>25.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>19/10</td>
<td>3.10e-08</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>$i$</td>
<td>59/35</td>
<td>4.85e-07</td>
<td>50.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>19/10</td>
<td>2.81e-06</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>$i$</td>
<td>44/21</td>
<td>4.20e-07</td>
<td>250.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/i$</td>
<td>19/10</td>
<td>1.56e-06</td>
<td>0.0136</td>
</tr>
</tbody>
</table>

Table 2: Results of RNA and Algorithm 2.1.
Moreover, we can see that for the same $\alpha$, $\beta$, $n$ and $x_0$, Table 2 and Table 3 show that the correction term does help to improve regularized Newton algorithm when the initial point is far away from the minimizer. These facts indicate that the introduction of correction is really useful and could accelerate the convergence of the regularized Newton method.

4. CONCLUSION
We propose a modified regularized Newton method for unconstrained nonconvex optimization, and show that if the gradient and Hessian of the objective function are Lipschitz continuous, then the M-RNM has a global convergence. The presented numerical results show that the practical of the proposed method. We may draw the conclusion that our method is better than the regularized Newton algorithm (RNM)[10] without correction.

5. REFERENCES

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>Dim</th>
<th>$x_0$</th>
<th>md</th>
<th>$|\nabla f|$</th>
<th>$x^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>10</td>
<td>(i)</td>
<td>14/6</td>
<td>1.43e-06</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>5/1</td>
<td>1.35e-09</td>
<td>0.2929</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>(i)</td>
<td>52/34</td>
<td>7.358e-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>6/2</td>
<td>1.16e-07</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>(i)</td>
<td>121/55</td>
<td>7.24e-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>5/1</td>
<td>4.15e-06</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>(i)</td>
<td>35/18</td>
<td>6.85e-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>6/2</td>
<td>5.89e-06</td>
<td>0.0136</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>10</td>
<td>(i)</td>
<td>5/1</td>
<td>1.45e-08</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>3/2</td>
<td>5.43e-08</td>
<td>0.2929</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>(i)</td>
<td>19/13</td>
<td>1.33e-08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>12/5</td>
<td>7.56e-07</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>(i)</td>
<td>39/24</td>
<td>4.28e-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>15/7</td>
<td>1.70e-06</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>(i)</td>
<td>43/26</td>
<td>9.80e-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>15/7</td>
<td>4.54e-06</td>
<td>0.0136</td>
</tr>
<tr>
<td>i</td>
<td>1</td>
<td>10</td>
<td>(i)</td>
<td>8/4</td>
<td>5.37e-08</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>11/4</td>
<td>3.79e-10</td>
<td>0.2929</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>(i)</td>
<td>33/16</td>
<td>4.20e-09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>17/10</td>
<td>3.13e-07</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>(i)</td>
<td>106/35</td>
<td>2.04e-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>28/19</td>
<td>1.28e-07</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>(i)</td>
<td>57/30</td>
<td>7.50e-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1/i)</td>
<td>38/21</td>
<td>4.17e-07</td>
<td>0.0136</td>
</tr>
</tbody>
</table>

Table 3: Results of RNA and Algorithm 2.1.


