Journal of Mathematical Extension
Vol. 12, No. 4, (2018), 115-135
ISSN: 1735-8299
URL: http://www.ijmex.com

A New Modified Trust Region Algorithm for Solving Unconstrained Optimization Problems

T. Dehghan Niri
Yazd University

M. M. Hosseini
Yazd University

M. Heydari*
Yazd University

Abstract. Iterative methods for optimization can be classified into two categories: line search methods and trust region methods. In this paper, we propose a modified regularized Newton method for minimizing nonconvex functions whose Hessian matrix may be singular without line search. The proposed method is proved to converge globally if the Gradient and Hessian of the objective function are Lipschitz continuous. Moreover, we report numerical results that show that the proposed algorithm is competitive with the existing methods.

AMS Subject Classification: 65K05; 90C26; 90C30
Keywords and Phrases: Regularized Newton method, unconstrained optimization, nonconvex, trust-region method, convergence analysis

1. Introduction

Unconstrained optimization problems have a number of important applications in many fields, such as operations research, economic equilibrium

Received: December 2017; Accepted: March 2018
*Corresponding author

models and engineering sciences. In these problems, the main goal is to find the minimum of the objective function with no restrictions at all on the values of variables. We consider the unconstrained optimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable and smooth function. Gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$ are denoted by $g(x)$ and $H(x)$, respectively. Throughout this paper, we assume that the solution set of (1) is nonempty.

There are many useful algorithms to solve unconstrained optimization problems such as: Newton method and modified Newton methods, quasi-Newton methods, conjugate gradient methods, trust region methods, etc. [3, 11, 14]. Among the methods mentioned above, the classical Newton method is very famous for its fast convergence property. There are several modifications of the Newton method for unconstrained minimization to achieve global and local convergence, see [3, 14] and the references therein. In Newton method, the positive definiteness of the Hessian matrix of the objective function is an essential condition to get the local minimum and the fast local convergence. At each iteration, the Newton method computes the trial step

$$d_k = -H_k^{-1} g_k, \tag{2}$$

where $g_k = g(x_k)$ and $H_k = H(x_k)$. To overcome the difficulty caused by the possible singularity of Hessian, Sun in [19] proposed a regularized Newton method, where the trial step is the solution of the linear equations

$$(H_k + \lambda_k I)d_k = -g_k, \tag{3}$$

where I is the identity matrix and λ_k is a positive parameter which is updated from iteration to iteration. Fan in [6] proposed $\lambda_k = \|g_k\|^\delta$ with $\delta \in [1, 2]$. Also Fan in [7, 8] showed that the choice of $\lambda_k = \|f_k\|$ performs more stable and preferable. A new trust region method for nonlinear equations with the trust region radius converging to zero is proposed in [5], and its convergence under some weak conditions is provided. Ueda and Yamashita [20] applied a regularized algorithm for nonconvex minimization problems. They gave a global complexity bound and analyzed the super linear convergence of their method. The disadvantage of this method is that calculating the most negative eigenvalue by decomposition methods or the method is computationally expensive. Also, in [21], they proposed a regularized Newton method without line search. Their method controls a regularized parameter instead of a step size in order to guarantee the global convergence. Wang in [22] proposed a modified regularized Newton method with correction for unconstrained nonconvex optimization. Also, he proved that the modified regularized Newton method has a global convergence and a local cubic convergence under some appropriate conditions. Shen et al. [17] proposed a regularized Newton method for solving unconstrained nonconvex minimization problems without the nonsingularity assumption of solutions. Under suitable conditions, the global convergence of the regularized Newton method and fast local convergence are established. Li in [12] showed that the regularized Newton method has quadratic convergence under the local error bound condition, where the trial step is the solution of the linear equations

$$(H_k + C\|g_k\|I)d_k = -g_k,$$

where C is a positive constant. Zhou in [23] proposed a two-step method for convex minimization problems whose Hessian matrices may be singular. Then solves the linear equations

$$(H_k + C\|g_k\|I)\widehat{d_k} = -g(y_k),$$

where $y_k = x_k + d_k$, to obtain the approximate Newton step $\widehat{d_k}$.

In this paper, we present an approximate of H_k and proposed a new algorithm for solving unconstrained nonconvex optimization and use $\lambda_k = \mu_k\|f_k\|$.

The organization of the paper is as follows: In Section 2, we present a new modified algorithms for solving nonconvex optimization problems. In Section 3, we show that the new algorithm preserves the same global convergence as the existing modified Levenberg-Marquardt (LM) algorithms under suitable conditions. The proposed method is tested on several examples taken from the literature and the numerical experiments are presented in Section 4. Finally, the conclusions are described in the last section.

2. The Algorithm

In this section, we introduce a regularized Newton method based on Zhou method [23]. We propose a new symmetric matrix instead of H_k and use a trust region technique to globalize the proposed method. Define the actual reduction of $f(x)$ at the $k-$th iteration as

$$Ared_k = f(x_k) - f(x_k + d_k + \widehat{d}_k). \tag{4}$$

We suggest that the regularized Newton step d_k is the minimizer of the problem:

$$\min_{d\in\mathbb{R}^n} \varphi_{k,1} = \frac{1}{2}d^T S_k d + g_k^T d + \frac{1}{2}\lambda_k\|d\|^2, \tag{5}$$

where S_k is a symmetric matrix with Hessian matrix properties. Let $S_k = (\frac{1}{f_k}g_k g_k^T - H_k)$, where $f_k = f(x_k)$ and $f_k \neq 0$. Also define

$$\Delta_{k,1} = \|d_k\| = \| - (B_k + \mu_k\|f_k\|I)^{-1} g_k^T f_k\|, \tag{6}$$

where $B_k = (g_k g_k^T - f_k H_k)$. Then similar to [18](Theorem 6.1.2), d_k is also a solution of the trust region problem:

$$\begin{aligned} &\min \frac{1}{2}d^T B_k d + g_k^T d, \\ &s.t. \qquad \|d_k\| \leqslant \Delta_{k,1}. \end{aligned}$$

Similar to the famous result given by Powell in [15], we know that

$$\varphi_{k,1}(0) - \varphi_{k,1}(d_k) \geqslant \frac{1}{2}\|f_k g_k\| \min\left\{\|d_k\|, \frac{\|f_k g_k\|}{\|B_k\|}\right\}. \tag{7}$$

Also, $\widehat{d}_k$ is the minimizer of the problem:

$$\min_{d\in\mathbb{R}^n} \varphi_{k,2} = \frac{1}{2}d^T B_k d + f(y_k)g_k^T d + \frac{1}{2}\lambda_k\|d\|^2, \tag{8}$$

and similar to d_k, $\widehat{d}_k$ is the solution of the following trust region problem:

$$\begin{aligned}\min_{d\in\mathbb{R}^n} \frac{1}{2}d^T B_k d + f(y_k)g_k^T d,\\ s.t. \qquad \|d_k\| \leqslant \Delta_{k,2},\end{aligned}$$

where

$$\Delta_{k,2} = \|\widehat{d}_k\| = \| - (B_k + \mu_k\|f_k\|I)^{-1}\, g_k^T f(y_k)\|. \tag{9}$$

Therefore, similar to (7)

$$\varphi_{k,2}(0) - \varphi_{k,2}(\widehat{d}_k) \geqslant \frac{1}{2}\|f(y_k)g_k\| \min\{\|\widehat{d}_k\|, \frac{\|f(y_k)g_k\|}{\|B_k\|}\}. \tag{10}$$

Now we define prediction reduction as

$$Pred_k = \varphi_{k,1}(0) - \varphi_{k,1}(d_k) + \varphi_{k,2}(0) - \varphi_{k,2}(\widehat{d}_k), \tag{11}$$

which satisfies

$$Pred_k \geqslant \frac{1}{2}\|f_k g_k\| \min\{\|d_k\|, \frac{\|f_k g_k\|}{\|B_k\|}\} + \frac{1}{2}\|f(y_k)g_k\| \min\{\|\widehat{d}_k\|, \frac{\|f(y_k)g_k\|}{\|B_k\|}\}. \tag{12}$$

The ratio of the actual reduction to the predicted reduction, $r_k = \frac{Ared_k}{Pred_k}$, plays an important role to decide that whether or not accept the trial step and how to adjust the regularized parameter. We set $\widehat{B}_k = B_k + E_k$ where $E_k = 0$ if B_k is positive definite, otherwise E_k is chosen to ensure that B_k is positive definite [14]. The regularized Newton algorithm for unconstrained optimization problems is stated as follows:

Algorithm 2.1 (Modified Regularized Newton Algorithm).

Input: $x_0 \in \mathbb{R}^n$, $\mu_1 > m > 0$, $0 < p_0 \leqslant p_1 \leqslant p_2 < 1$ and $\epsilon > 0$.

Step2. If $\|f_k g_k^T\| = 0$, then stop.

Solve

$$(\widehat{B}_k + \mu_k \|f_k\| I) d_k = -f_k g_k^T.$$

Set

$$y_k = x_k + d_k.$$

Solve

$$(\widehat{B}_k + \mu_k \|f_k\| I) \widehat{d}_k = -f(y_k) g_k^T.$$

Set

$$s_k = d_k + \widehat{d}_k.$$

Step3. Compute $r_k = \frac{Ared_k}{Pred_k}$. Set

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{if } \; r_k \geqslant p_0, \\ x_k, & \text{otherwise.} \end{cases}$$

Step4. Update μ_{k+1} as

$$\mu_{k+1} = \begin{cases} 4\mu_k, & \text{if } \; r_k < p_1, \\ \mu_k, & \text{if } \; r_k \in [p_1, p_2], \\ \max\{\frac{\mu_k}{4}, m\}, & \text{if } \; r_k > p_2. \end{cases}$$

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

Lemma 2.1. *Let the sequence* $\{x_k\}$ *is generated by Algorithm* 2.1*, then the sequence* $\{f(x_k)\}$ *is decreasing.*

Proof. If $r_k < p_0$, according to Algorithm 2.1, $x_{k+1} = x_k$ and so $f(x_{k+1}) = f(x_k)$. So, we can let $r_k \geqslant p_0 > 0$.
Then (12) implies that $Pred_k \geqslant 0$ and according to the definition of r_k, we can say that $Ared_k > 0$. Therefore by (4), we have $f(x_k) - f(x_k + d_k + \widehat{d}_k) > 0$. Then $f(x_k) > f(x_k + d_k + \widehat{d}_k)$ and so, the sequence $\{f(x_k)\}$ is a decreasing sequence. $\square$

3. Global Convergence

In this section, we study the global convergence of Algorithm 2.1. We first give the following assumptions.

Assumption 3.1. $f(x)$, $g(x)$ and $H(x)$ are Lipschitz continuous, that is, there exists positive constants L_1, L_2 and L_3 such that

$$\|f(y) - f(x)\| \leqslant L_1\|y - x\|, \quad x, y \in \mathbb{R}^n, \tag{13}$$

$$\|g(y) - g(x)\| \leqslant L_2\|y - x\|, \quad x, y \in \mathbb{R}^n, \tag{14}$$

and

$$\|H(y) - H(x)\| \leqslant L_3\|y - x\|, \quad x, y \in \mathbb{R}^n. \tag{15}$$

Without loss of generality, suppose $L = \max(L_1, L_2, L_3)$.

Assumption 3.2. The mapping f is twice continuously differentiable and the level set

$$L(x_0) = \{x \in \mathbb{R}^n | f(x) \leqslant f(x_0)\},$$

is bounded.

Lemma 3.3. *Suppose A is symmetric positive semidefinite. Then,*

$$\|A + \varphi I\| \geqslant \varphi, \tag{16}$$

and

$$\|(A + \varphi I)^{-1}\| \leqslant \varphi^{-1}, \tag{17}$$

hold for any $\varphi > 0$.

Proof. See [8]. □

Theorem 3.4. *Let Assumptions 3.1 and 3.2 hold. Then Algorithm 2.1 terminates in finite iterations or satisfies*

$$\lim_{k\to\infty} \|g_k^T f_k\| = 0. \tag{18}$$

Proof. We use the contradiction to prove the theorem in a similar manner with [9]. Suppose (18) is not true, then there exists $\epsilon > 0$ and an integer $\widehat{k}$ such that

$$\|g_k^T f_k\| \geqslant \epsilon, \qquad \forall k \geqslant \widehat{k}. \tag{19}$$

Without loss of generality, suppose $\widehat{k} = 1$. Set $T = \{k | x_{k+1} \neq x_k\}$, Then

$$\{1, 2, \ldots\} = T \cup \{k | x_{k+1} = x_k\}.$$

Now we will analysis the following two cases:

Case (a): Suppose T is finite. Then there exists an integer k_1 such that

$$x_{k_1} = x_{k_1+1} = x_{k_1+2} = \ldots.$$

Therefore, according to Step 3 of Algorithm 2.1, we have

$$r_k < p_0, \qquad \forall k \geqslant k_1.$$

Therefore by Step 4 of Algorithm 2.1, we deduce

$$\mu_k \to \infty, \qquad \lambda_k \to \infty, \tag{20}$$

where $\lambda_k = \mu_k \|f_k\|$. Since $x_{k+1} = x_k$, $\forall k \geqslant k_1$, from relation (20) and definition of d_k in Algorithm 2.1 and Lemma 3.3 we get

$$\|d_k\| = \| - (\widehat{B}_k + \lambda_k I)^{-1} f_k g_k\| \leqslant \mu_k^{-1} \|g_k\| \to 0. \tag{21}$$

From the definition of $\widehat{d_k}$ in Algorithm 2.1 and by using (13), (20) and assuming that $g(x^*) = 0$, we have

$$\|\widehat{d_k}\| = \| - (\widehat{B}_k + \lambda_k I)^{-1} f(y_k) g_k^T\|$$

$$\leqslant \|(\widehat{B}_k + \lambda_k I)^{-1}(f(y_k) - f_k) g_k^T\| + \|(\widehat{B}_k + \lambda_k I)^{-1} f_k g_k^T\|$$

$$\leqslant L\|d_k\|\|(\widehat{B}_k + \lambda_k I)^{-1} g_k^T\| + \|d_k\|$$

$$\leqslant \Big(\frac{C_1}{\lambda_k}\|x_k - x^*\|^2 + 1\Big)\|d_k\| \leqslant K\|d_k\|, \tag{22}$$

where C_1 and K are positive constants. According to definition of (4) and (11), we have

$$|Ared_k - Pred_k| =$$

$$\Big|\Big(f(x_k) - f(x_k + d_k + \widehat{d}_k)\Big) - \Big(\varphi_{k,1}(0) - \varphi_{k,1}(d_k) + \varphi_{k,2}(0) - \varphi_{k,2}(\widehat{d}_k)\Big)\Big|$$

$$\leqslant \Big|f(y_k + \widehat{d}_k) - f(y_k) - \frac{1}{2}\widehat{d}_k^T \widehat{B}_k \widehat{d}_k - f(y_k) g_k^T \widehat{d}_k\Big|$$

$$+\Big|f(y_k) - f(x_k) - \frac{1}{2} d_k^T \widehat{B}_k d_k - f_k g_k^T d_k\Big|$$

$$= o(\|d_k\|^2) + o(\|\widehat{d}_k\|^2). \tag{23}$$

Moreover, from (12), (14), (19) and (21), we have

$$Pred_k \geqslant \frac{1}{2}\tau \min\Big\{\|d_k\|, \frac{\tau}{L}\Big\} \geqslant \frac{1}{2}\tau\|d_k\|, \tag{24}$$

for sufficiently large k. According to (23) and (24), we get

$$\Big|r_k - 1\Big| = \Big|\frac{Ared_k - Pred_k}{Pred_k}\Big| = \frac{o(\|d_k\|^2) + o(\|\widehat{d}_k\|^2)}{\|d_k\|} \to 0, \tag{25}$$

which implies that $r_k \to 1$. Therefore from Step 4 in Algorithm 2.1, there exists constant $\xi > 0$ such that

$$\mu_k \leqslant \xi,$$

which contradicts to the basic assumption (19).

Case b: Suppose T is an infinite set. From Assumption 3.2, (12) and (19), we have

$\infty > f(x_1) - \liminf_{k\to\infty} f(x_k) \geqslant \sum_{i=1}^{\infty} f(x_i) - f(x_{i+1})$

$= \sum_{k\in T} f(x_k) - f(x_{k+1}) \geqslant \sum_{k\in T} p_0 Pred_k$

$\geqslant p_0 \left(\frac{1}{2}\|f_k g_k\| \min\left\{\|d_k\|, \frac{\|f_k g_k\|}{\|\widehat{B}_k\|}\right\} + \frac{1}{2}\|f(y_k) g_k\| \min\left\{\|\widehat{d}_k\|, \frac{\|f(y_k) g_k\|}{\|\widehat{B}_k\|}\right\}\right)$

$$\geqslant \sum_{k\in T} p_0 \frac{\tau}{2} \min\left\{\|d_k\|, \frac{\tau}{L}\right\}, \tag{26}$$

which relation (26) implies that

$$\lim_{k\to\infty, k\in T} d_k = 0. \tag{27}$$

From (27) and μ_k produced by Algorithm 2.1, we have

$$\lambda_k \to \infty. \tag{28}$$

$$\|s_k\| = \|d_k + \widehat{d}_k\| \leqslant \|d_k\| + \|\widehat{d}_k\| \leqslant c\|d_k\|, \qquad \forall\ k \in T. \tag{29}$$

This equality together with (26) yields

$$\sum_{k\in T} \|s_k\| = \sum_{k\in T} \|d_k + \widehat{d}_k\| < \infty, \tag{30}$$

which implies that

$$\sum_{k\in T} \|x_{k+1} - x_k\| < \infty. \tag{31}$$

Then

$$x_k \to \overline{x}. \tag{32}$$

From definition of d_k, (22), (28) and (32), we get

$$d_k \to 0, \qquad \widehat{d}_k \to 0. \tag{33}$$

Since $\left(\widehat{B}_k + \lambda_k I\right) d_k = -g_k^T f_k$ then from (19), we have

$$\lambda_k \|d_k\| = \|g_k^T f_k + \widehat{B}_k d_k\| \geqslant \|g_k^T f_k\| - \|\widehat{B}_k\| \|d_k\| \geqslant \tau - \|\widehat{B}_k\| \|d_k\|, \quad (34)$$

therefore from (13), (14) and (15)

$$\lambda_k \geqslant \frac{\tau}{\|d_k\|} - \|\widehat{B}_k\| \geqslant \frac{\tau}{\|d_k\|} - \|g_k\|^2 + \|f_k\| \|H_k\| \geqslant \frac{\tau}{\|d_k\|} + C, \quad (35)$$

where C is a positive constant. Which (35) according to (33) means

$$\lambda_k \to \infty. \quad (36)$$

By the same analysis as (25) we know that $r_k \to 1$. Hence, there exists a positive constant $\eta > m$ such that $\mu_k \leqslant \eta$ holds for all sufficiently large k, which implies a contradiction to (19). Therefore initial assumption is false and the proof is completed. $\square$

4. Local Convergence

In this section, we study the local convergence properties of the proposed algorithm. In a similar manner with [16], the local convergence theory requires the following assumptions. We assume that the solution set of (1) is nonempty and denote it by X^*. Also $\{x_k\}$ converges to $x^* \in X^*$ and lies in some neighborhood of x^*.

Assumption 4.1.
(I) There exists a solution $x^* \in X^*$ of (1).
(II) $g(x)$ is Lipschitz continuous on $N(x^*, b) = \{x \in \mathbb{R}^n | \ \|x - x^*\| \leqslant b\}$, i.e., there exists a positive constant L such that

$$\|g(y) - g(x)\| \leqslant L\|y - x\|, \quad \forall x, y \in N(x^*, b), \quad (37)$$

where $0 < b < 1$.

Assumption 4.2. (**A**) $\|g(x)\|$ provides a local error bound on some neighbourhood of x^*, i.e., there exist two positive constants c and b such that

$$\|g(x)\| \geqslant c \ dist(x, X^*), \qquad \forall x \in N(x^*, b), \quad (38)$$

(**B**) The Hessian $H(x)$ is Lipschitz continuous on $N(x^*, b)$, that is, there exists a constant L such that

$$\|H(y) - H(x)\| \leqslant L\|y - x\|, \quad \forall x, y \in N(x^*, b). \tag{39}$$

Lemma 4.3. According to Assumptions 3.2, 4.1 (II), 4.2 (B) and definition of S, there exist constants K_1 and C_1 such that

$$\|S(y) - S(x)\| \leqslant C_1\|y - x\|^2 + K_1\|y - x\|, \quad \forall x, y \in N(x^*, b). \tag{40}$$

Proof. Under Assumptions 4.1 (II) and 4.2 (B), and since $\{f(x_k)\}$ is a monotone decreasing sequence and has a bound from below, we have

$$\|S(y) - S(x)\| \leqslant M\|g(y)g(y)^T - g(x)g(x)^T\| + \|H(y) - H(y)\| \leqslant$$

$$M\left(\|g(y)g(y)^T - g(y)g(x)^T\| + \|g(y)g(x)^T - g(x)g(x)^T\|\right) + L\|y - x\| \leqslant$$

$$M\left(\|g(y)\|\|g(y)^T - g(x)^T\| + \|g(y) - g(x)\|\|g(x)^T\|\right) + L\|y - x\| \leqslant$$

$$ML\|y - x\| \left(\|g(y)\| + \|g(x)\|\right) + L\|y - x\| \leqslant$$

$$ML\|y - x\|\Big(\|g(y) - g(x)\| + 2\|g(x) - g(x^*)\|\Big) + L\|y - x\| \leqslant$$

$$C_1\|y - x\|^2 + K_1\|y - x\|,$$

where $C_1 = ML^2$ and $K_1 = L(1 + 2MLb)$. $\square$

In the following, we denote $\overline{x}_k$ the vector in the solution set X^* that satisfies

$$dist(x_k, X^*) = \|x_k - \overline{x}_k\|.$$

To obtain faster convergence of the modified regularized Newton method, we need to estimate $\|\widehat{d}_k\|$ more accurately. We will use the SVD technique to derive the local convergence rate of algorithms 2.1. Since $\widehat{B}_k(x^*)$ is a symmetric matrix, there is an orthogonal matrix (U_1^*, U_2^*) such that

$$\widehat{B}_k(x^*) = (U_1^*, U_2^*) \begin{pmatrix} \Sigma_1^* & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} {V_1^*}^T \\ {V_2^*}^T \end{pmatrix},$$

where Σ_1^* is a diagonal matrix. Also, we can suppose that $\widehat{B}_k(x)$ has the following decomposition,

$$\widehat{B}_k(x) = (U_1, U_2) \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T,$$

where $rank(\Sigma_1) = rank(\Sigma_1^*)$ and Σ_2 converges to zero as $x \to x^*$. We first prove the linear convergence of Algorithm 2.1, which implies that $\|x_k - x^*\| = dist(x_k, X^*)$.

Lemma 4.4. *Under Assumption 4.1, if $x_k, y_k \in N(x^*, \frac{b}{2})$, we have*

$(a) \|d_k\| \leqslant c_1 \; dist(x_k, X^*)$,
$(b) \|\widehat{d}_k\| \leqslant c_2 \; dist(x_k, X^*)$,
$(c) \|s_k\| \leqslant c_3 \; dist(x_k, X^*)$,

for sufficiently large k, where c_1, c_2 and c_3 are positive constants.

Proof. The proof is similar to Lemma 3.2 in [10]. □

Lemma 4.5. *Under Assumption 4.1, if $x_k, y_k \in N(x^*, \frac{b}{2})$, then there exists a positive constant $\delta > m$ such that for all sufficiently large k,*

$$\mu_k \leqslant \delta, \tag{41}$$

holds.

Proof. Proof in [10]. □

Lemma 4.6. *Let Assumptions 4.1 and 4.2 hold. Then we have*

$$dist(x_{k+1}, X^*) = O(dist(x_k, X^*)). \tag{42}$$

Proof. According to Assumptions 4.1, 4.2 and Lemma 4.3, we have

$$c \; \|\overline{x}_{k+1} - x_{k+1}\| \leqslant \|g(x_{k+1})\| = \|g(y_k + \widehat{d}_k)\|$$

$$\leqslant \|g(y_k + \widehat{d}_k) - g(y_k) - S(y_k)\widehat{d}_k\| + \|g(y_k) + S(y_k)\widehat{d}_k\|$$

$$\leqslant L\|\widehat{d}_k\|^2 + \|g(y_k) + S_k\widehat{d}_k\| + \|S(y_k)\widehat{d}_k - S_k\widehat{d}_k\|$$

$$\leqslant L\|\widehat{d}_k\|^2 + \|g(y_k) + S_k\widehat{d}_k\| + \Big(C_1\|d_k\|^2 + K_1\|d_k\|\Big)\|\widehat{d}_k\|$$

$$\leqslant L\|\widehat{d}_k\|^2 + \Big(C_2\|d_k\| + \lambda_k\|\widehat{d}_k\|\Big) + \Big(C_1\|d_k\|^2 + K_1\|d_k\|\Big)\|\widehat{d}_k\|.$$

Therfore from Lemma 4.4 and $\lambda_k = \mu_k\|f_k\|$, we get

$$c\ \|\overline{x}_{k+1} - x_{k+1}\| \leqslant \|g(x_{k+1})\| \leqslant k_1\|d_k\| + k_2\|d_k\|^2 + k_3\|d_k\|^3,$$

where k_1, k_2 and k_3 are positive constants. □

Theorem 4.7. *Let Assumption 4.2 hold. Then we have*

$$\|s_{k+1}\| = O(\|s_k\|), \qquad \|x_{k+1} - x^*\| = O(\|x_k - x^*\|). \tag{43}$$

Proof. Proof in [23]. □

Lemma 4.8. *Under Assumption 4.1, if $x_k, y_k \in N(x^*, \frac{b}{2})$, then we have*

(a) $\|g(y_k)\| \leqslant O(\|\overline{x}_k - x_k^*\|)$,

(b) $\|U_2U_2^T g(y_k)\| \leqslant O(\|\overline{x}_k - x_k^*\|^2)$,

for all sufficiently large k.

Proof. We can find the proofs of (a) and (b) in [23]. □

Theorem 4.9. *Let the sequence $\{x_k\}$ is generated by Algorithm 2.1, under the conditions of Assumption 4.1 the sequence $\{x_k\}$ converges quadratically to a solution of (1).*

Proof. Theorem is proved in a similar manner with [9] and [23]. □

5. Numerical Results

In this section, we report some results on the following numerical experiments for the proposed algorithm (Algorithm 2.1). Also compare the effectiveness of the proposed method with the extended regularized Newton method (E-RN method) [20], regularized Newton method with correction [22] and Modified cholesky method [14]. In Algorithm 2.1 and RN method, suppose $p_0 = 0.0001, p_1 = 0.25, p_2 = 0.75, \mu_1 = 10^{-5}, m = 10^{-8}$, also in E-RN method $c_1 = 2$, $c_2 = 10^{-5}, \alpha = 10^{-4}$. The stopping criterion is $\|g(x_k)\| \leqslant 10^{-5}$. N_f represents the number of evaluations of the objective function, N_g represents the number of evaluations of its gradient and "Dim" shows the dimension of the problem. All of the algorithms are implemented in Matlab 12.0 and runs are made on 2.3 GHz PC with 8 GB memory. The test functions commonly used unconstrained test problems with standard starting points (see [1, 2, 13]) and summary of which is given in Table 1.

Table 1: Test Problems

No.	Name	No.	Name
1	Example 1	12	TRIDIA
2	Example 2	13	Diagonal Double Bounded Arrow Up
3	Extended Beale	14	NONDIA
4	Extended White-Holst	15	FLETCHCR
5	Brown badly scaled	16	DENSCHNB
6	Extended Powell Singular	17	DENSCHNF
7	Freudenstein and Roth	18	DIXON3DQ
8	Extended Tridiagonal 1	19	BIGGSB1
9	Extended DENSCHNB	20	DIAG-AUP1
10	Extended Himmelblau	21	Griewank
11	NONDQUAR	22	Broyden Tridiagonal

Example 1. [17] $f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_1^2x_2^2, x_0 = (-1.2, 1), f(x^*) = 0.$

Table 2: Numerical results

NO./Dim	Proposed Method N_f/N_g CPU time(s) $\|f_k - f^*\|$	RN method [22] N_f/N_g CPU time(s) $\|f_k - f^*\|$	E-RN method [20] N_f/N_g CPU time(s) $\|f_k - f^*\|$	Modified Cholesky [14] N_f/N_g CPU time(s) $\|f_k - f^*\|$
1/2	14/8 0.867 1.26 E-13	26/27 1.615 0	13/14 1.475 1.92 E-12	25/25 0.705 3.02 E-44
2/300	12/7 6.752 1.80 E-13	28/29 17.172 0	13/13 13.984 1.74 E-12	20/19 24.077 6.47 E-28
3/10	10/6 0.366 6.55 E-17	28/29 2.460 1.04 E-11	23/23 3.268 7.78 E-12	44/9 3.010 1.29 E-13
3/300	10/6 9.756 1.05 E-7	FAIL - -	25/25 47.476 1.47 E-11	45/10 27.188 3.99 E-24
4/200	42/22 12.063 3.12 E-21	FAIL - -	51/50 31.046 8.15 E-11	11/10 5.277 2.47 E-13
5/2	22/12 0.382 7.93 E-27	40/41 2.709 0	22/23 0.704 0	FAIL - -
6/4	12/7 0.226 5.97 E-17	22/23 1.512 2.48 E-9	15/16 0.580 4.40 E-9	17/18 1.915 1.71 E-10
6/96	22/12 3.013 3.51 E-17	24/25 4.294 6.22 E-9	16/17 5.380 2.08 E-8	18/19 5.626 8.10 E-10
7/10	122/62 4.134 1.04 E-28	10/11 0.886 1.96 E+2	6/7 0.916 1.96 E+2	6/7 0.913 1.96 E+2
8/10	12/7 0.286 1.73 E-10	18/19 1.364 7.13 E-9	12/13 1.480 1.76 E-8	14/15 1.712 6.88 E-10
8/300	12/6 6.431 5.53 E-9	20/21 11.388 2.22 E-8	13/14 14.715 1.05 E-7	15/16 15.150 4.08 E-9
9/300	14/8 7.785 1.42 E-20	34/35 19.262 1.48 E-29	35/26 27.870 6.59 E-16	30/4 7.500 0
10/300	12/7 6.648 6.84 E-27	54/55 32.854 8.29 E-16	136/134 159.513 8.59 E-13	6/7 6.805 6.20 E-16
11/10	12/7 0.307 4.18 E-20	18/19 1.279 1.14 E-8	13/14 1.699 5.58 E-9	15/16 1.867 2.18 E-10

Example 2. [17] $f(x) = \sum_{i=1}^{n} (\frac{1}{2}(x_{3i-2} - \frac{x_{3i-1}}{2})^2 + \frac{1}{2}(x_{3i-2} - \frac{x_{3i-1}}{2})^2 x_{3i}^2)$, $f(x^*) = 0$, $x_0 = (-1.2, 1, \ldots, -1.2, 1)$.

Table 3: Numerical results

NO./Dim	Proposed Method N_f/N_g CPU time(s) $\|f_k - f^*\|$	RN method [22] N_f/N_g CPU time(s) $\|f_k - f^*\|$	E-RN method [20] N_f/N_g CPU time(s) $\|f_k - f^*\|$	Modified Cholesky [14] N_f/N_g CPU time(s) $\|f_k - f^*\|$
12/10	4/3 0.265 1.71 E-28	2/3 0.145 5.62 E-12	2/3 0.258 4.83 E-21	1/2 0.186 1.50 E-29
13/10	14/8 0.346 6.58 E-21	FAIL - -	10/11 1.285 7.25 E-20	28/12 2.123 4.93 E-15
13/100	38/20 6.647 9.05 E-17	FAIL - -	134/132 44.304 5.74 E-13	11/10 3.226 8.90 E-20
14/10	16/9 1.058 1.49 E-14	68/69 4.047 1.24 E-28	67/66 8.083 9.90 E-1	15/16 4.956 9.90 E-1
14/300	12/7 6.476 1.90 E-17	8/9 4.566 7.41 E-24	7/8 7.873 1.38 E-22	7/8 7.956 5.61 E-24
15/10	14/8 0.382 2.00 E-27	FAIL - -	36/36 3.380 1.31 E-14	17/7 1.397 1.72 E-23
16/10	36/19 2.353 1.08 E-12	34/35 2.885 2.27 E-12	19/20 2.244 3.09 E-17	141/34 10.194 1.15 E-14
17/10	10/6 0.356 1.09 E-21	8/9 0.923 6.72 E-19	6/7 0.983 3.26 E-21	6/7 0.976 2.26 E-21
18/10	4/3 0.259 5.48 E-29	2/3 0.146 4.23 E-14	2/3 0.256 1.69 E-22	1/2 0.221 1.97 E-31
19/10	4/3 0.262 1.51 E-17	2/3 0.146 6.61 E-16	2/3 0.270 6.61 E-25	1/2 0.192 4.93 E-32
20/200	14/8 5.314 2.31 E-19	14/15 5.490 1.46 E-26	9/10 6.708 4.25 E-15	10/11 7.258 7.44 E-26
21/10	8/5 1.194 2.65 E-14	FAIL - -	17/17 4.408 1.55 E-10	28/7 3.091 0
21/50	8/5 67.560 5.06 E-14	FAIL - -	31/30 501.544 1.54 E-10	72/10 189.557 0
22/300	8/5 8.210 1.15 E-12	8/9 9.116 5.49 E-15	6/7 12.347 3.34 E-17	6/7 12.362 3.34 E-17

The results on above problems are listed in Tables 2 and 3. The proposed method has considerable advantage in number of evaluations, the error and computational time. Recently, for comparison of iterative algorithms, Dolan and More' [4] proposed a new technique comparing the considered algorithms with statistical process by demonstration of performance profiles. In this process it is known that a plot of the performance profile reveals all of the major performance characteristics, which is a common tool to graphically compare effectiveness and robustness of the algorithms. In this technique, one can choose a performance index as measure of comparison among considered algorithms and can illustrate the results with performance profile. Figures 1 and 2 show the comparisons of proposed method (Algorithm 2.1), RN method [22], E-RN method [20] and Modified Cholesky method [14] relative to computing time, the number of evaluations of the objective function (N_f) and the number of evaluations of its gradient (N_g), respectively.

6. Conclusions

In this paper, we propose a new modified Newton method for unconstrained minimization problems and analyze its global and local convergence. Using this algorithm, convex and nonconvex problems can be solved. We also test our algorithm on some unconstrained problems. The numerical results and comparison with some algorithms confirm the efficiency and robustness of our algorithm. Finally, we give detailed computational experiments and numerical comparisons to show that our approach is potentially efficient.

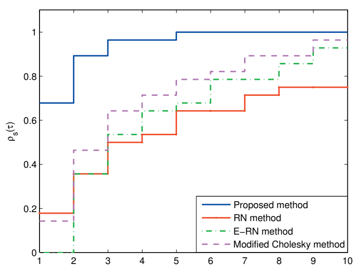

Figure 1. Performance profile for CPU time

(a) (b)

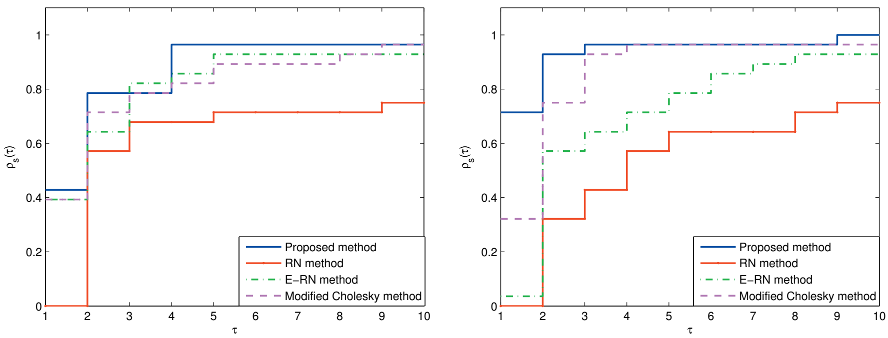

Figure 2. (a) Performance profile for N_f
(b)Performance profile for N_g

Acknowledgements

The authors are grateful for the valuable comments and suggestions of referees, which improved this paper.

References

[1] N. Andrei, *Test functions for unconstrained optimization,* 8-10, Averescu Avenue, Sector 1, Bucharest, Romania. Academy of Romanian Scientists (2004).

[2] N. Andrei, An unconstrained optimization test functions collection, *Adv. Model. Optim.,* 10 (2008), 147-161.

[3] J. E. Dennis and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equation*, (1996).

[4] E. D. Dolan and J. J. More, Benchmarking optimization software with performance profiles, Math. *Program.,* 91 (2002), 201-213.

[5] J. Y. Fan, Convergence rate of the trust region method for nonlinear equations under local error bound condition. *Comput. Optim. Appl.,* 34 (2006), 215-227.

[6] J. Y. Fan and Y. X. Yuan, On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. *Computing,* 74 (2005), 23-39.

[7] J. Y. Fan and J. Y. Pan, A note on the Levenberge-Marquardt parameter. *Appl. Math. Comput.,* 207 (2009), 351-359.

[8] J. Y. Fan and Y. X. Yuan, A regularized Newton method for monotone nonlinear equations and its application. *Optim. Methods Softw.,* 29 (2014), 102-119.

[9] J. Y. Fan, The modified Levenberg-Marquardt method for nonlinear equations with cubic convergence, *Math. Comp.,* 81 (2012), 447-466.

[10] J. Fan, Accelerating the modified Levenberg-Marquardt method for nonlinear equations. *Math. Comp.,* 83 (2014), 1173-1187.

[11] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations,* North Carolina State University (1995).

[12] D. H. Li, M. Fukushima, L. Qi, and N. Yamashita, Regularized Newton methods for convex minimization problems with singular solutions, *Comput. Optim. Appl.,* 28 (2004), 131-147.

[13] J. J. More, B. S. Grabow, and K. E. Hillstrom, testing unconstrained optimization software, *ACM, Trans. Math. software*, 7 (1981), 17-41.

[14] J. Nocedal and S. Wright, *Numerical Optimization,* 2nd edn. Springer, New York (2006).

[15] M. J. D. Powell, *Nonlinear Programming. In: O.L. Mangasarian, R. R. Meyer, S. M. Robinson (Eds.)* Convergence properties of a class of minimization algorithms, 2, 1-27. Academic Press, New York. (1975).

[16] L. Qi, Xi. Xiao, and Li. Zhang, On the global convergence of a parameter-adjusting levenberg-marquardt method, *Num. Alg. Control and Optimization.,* 5 (2015), 25-36.

[17] Ch. Shen, Xi. Chen, and y. Liang, A regularized Newton method for degenerate unconstrained optimization problems, *Optim. Lett.,* 6 (2012), 1913-1933.

[18] W. Sun and Y. Yuan, *Optimization Theory and Methods,* Springer Science and Business Media, LLC, New York, (2006).

[19] D. Sun, A regularization Newton method for solving nonlinear complementarity problems, *Appl. Math. Optim.,* 40 (1999), 315-339.

[20] K. Ueda and N. Yamashita, Convergence properties of the regularized Newton method for the unconstrained nonconvex optimization. *Appl. Math. Optim.,* 62 (2010), 27-46.

[21] K. Ueda and N. Yamashita, A regularized Newton method without line search for unconstrained optimization, *Comput. Optim. Appl.,* 59 (2014), 321-351.

[22] H. Wang and M. Qin, A Regularized Newton Method with Correction for Unconstrained Nonconvex Optimization, *J. Math. Res.,* 7 (2015), 7-17.

[23] W. Zhou and X. Chen, On the convergence of a modified regularized Newton method for convex optimization with singular solutions, *J. Comput. Appl. Math.,* 239 (2013), 179-188.

Tayebeh Dehghan Niri
Ph.D Student of Mathematics
Department of Mathematics
Yazd University
Yazd, Iran
E-mail: T.dehghan@stu.yazd.ac.ir

Mohammad Mehdi Hosseini
Professor of Mathematics
Department of Mathematics
Yazd University
Yazd, Iran
E-mail: hosse_m@yazd.ac.ir

Mohammad Heydari
Associate Professor of Mathematics
Department of Mathematics
Yazd University
Yazd, Iran
E-mail: m.heydari@yazd.ac.ir