A Modified Regularized Newton Method for Unconstrained Convex Optimization

Heng Wang , Mei Qin*

Abstract—**In this paper, we present a modified regularized Newton method (M-RNM) for minimizing a convex function whose Hessian matrices may be singular. At every iteration, not only a RNM step is computed but also two correction steps are computed. We show that if the objective function is LC^2, then the method posses local quadratic convergence under a local error bound condition. A globally convergent M-RNM algorithm is also given by using trust region technique.**

Index Terms—**Regularized Newton method, Local error bound, Correction technique, Trust region technique, Unconstrained convex optimization.**

I. Introduction

We consider the unconstrained optimization problem [1-5]

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

where $f : R^n \to R$ is a convex and twice continuously differentiable, whose 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 and denoted by X, and in all cases $\|\cdot\|$ 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 $x \in R^n$. Moreover, if $f(x)$ is convex, then $x \in X$ if and only if x is a solution of the following nonlinear equations

$$g(x) = 0. \tag{2}$$

Hence, we could get the minimizer of $f(x)$ by solving (2) [6-9].The Newton method is one of the efficient solution method. At every iteration, it computes the trial step

$$d_k^N = -H_k^{-1} g_k, \tag{3}$$

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. In this case, we may compute the Moore-Penrose step $d_k^{MP} = -H_k^{+} g_k$. But the computation of the singular value decomposition to obtain H_k^{+} is sometimes prohibitive. Hence, computing a direction that is close to d_k^{MP} may be a good idea.

To overcome the difficulty caused by the possible singularity of H_k, [10] proposed a regularized Newton method, where the trial step is the solution of the linear equations

$$(H_k + \lambda_k I)d = -g_k, \tag{4}$$

Manuscript received July 24, 2015; revised October 19, 2015. This work was supported by the National Natural Science Foundation of China(11426155) and the Hujiang Foundation of China (B14005).

Heng Wang is with the College of Science, University of Shanghai for Science and Technology, Shanghai. e-mail:wanghengusst@126.com

Mei Qin is with the College of Science, University of Shanghai for Science and Technology, Shanghai, 200093, P.R. China. (*Corresponding Author: E-mail Address: usstjssx@126.com)

where I is the identity matrix. λ_k is a positive parameter which is updated from iteration to iteration.

Now we need to consider another question, "how to choose the regularized parameter λ_k ? " Yamashita and Fukushima [11] 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 [12] took $\lambda_k = \|g_k\|^\delta$ with $\delta \in [1, 2]$ and showed that the Levenberg-Marquardt method preserves the quadratic convergence. Numerical results ([13], [14]) show that the choice of $\lambda_k = \|F_k\|$ performs more stable and preferable.

Inspired by the regularized Newton method [14] with correction for nonlinear equations, we propose a modified regularized Newton method in this paper. At every iteration, the modified regularized Newton method solves the linear equations

$$(H_k + \lambda_k I)d = -g_k \quad with \ \ \lambda_k = \mu_k \|g_k\| \tag{5}$$

to obtain the Newton step d_k, where $\mu_k > 0$ is updated from iteration to iteration, and then solves the linear equations

$$(H_k + \lambda_k I)\, d = -g_k + \lambda_k d_k \tag{6}$$

to obtain the approximate Newton step s_k.

It is easy to see

$$s_k = d_k + \widetilde{d_k}, \quad \widetilde{d_k} = \lambda_k (H_k + \lambda_k I)^{-1} d_k. \tag{7}$$

Finally, the M-RNM solves the linear equations

$$(H_k + \lambda_k I)\, s = -g\,(y_k) \quad with\ y_k = x_k + s_k \tag{8}$$

to obtain the approximate Newton step $\widetilde{s_k}$.

The aim of this paper is to study the convergence properties of the modified regularized Newton method.

The paper is organized as follows. In Section 2, we present a new modified regularized Newton algorithm by using trust region ([14], [21], [22], [24]) technique, and prove the global convergence of the new algorithm under some suitable conditions. In Section 3, we study the convergence rate of the algorithm, and obtain the quadratic convergence under the local error bound condition. Finally, we conclude the paper in Section 4.

II. The algorithm and its global convergence

First, we give the modified regularized Newton algorithm. Define the actual reduction of $f(x)$ at the kth iteration as

$$Ared_k = f(x_k) - f(x_k + s_k + \widetilde{s_k}). \tag{9}$$

Note that the regularization step d_k is the minimizer of the convex minimization problem

$$\min_{d\in R^n} \frac{1}{2} d^T H_k d + g_k^T d + \frac{1}{2}\lambda_k \|d\|^2.$$

(Advance online publication: 14 May 2016)

If we let

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

then it can be proved [6] that d_k is also a solution of the trust region problem

$$\min_{d \in R^n} \varphi(d) = \frac{1}{2} d^T H_k d + g_k^T d, \ s.t.\ \|d\| \le \Delta_{k,1}.$$

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

$$\varphi(0) - \varphi(d_k) \ge \frac{1}{2} \|g_k\| \min\left\{ \|d_k\|, \frac{\|g_k\|}{\|H_k\|} \right\}. \quad (10)$$

By some simple calculations, we deduce from (7) that

$$\begin{aligned} \varphi(d_k) - \varphi(s_k) &= g_k^T d_k + \frac{1}{2} d_k^T H_k d_k - g_k^T s_k - \frac{1}{2} s_k^T H_k s_k \\ &= -g_k^T \widetilde{d_k} - \frac{1}{2} \widetilde{d_k}^T H_k \widetilde{d_k} - \widetilde{d_k}^T H_k d_k \\ &= \lambda_k \widetilde{d_k}^T d_k - \frac{1}{2} \widetilde{d_k}^T H_k \widetilde{d_k} \\ &= \frac{1}{2} \widetilde{d_k}^T H_k \widetilde{d_k} + \lambda_k \widetilde{d_k}^T \widetilde{d_k} \\ &\ge 0, \end{aligned}$$

so, we have

$$\varphi(0) - \varphi(s_k) \ge \varphi(0) - \varphi(d_k). \quad (11)$$

Similar to d_k, $\widetilde{s_k}$ is not only the minimizer of the problem

$$\min_{s \in R^n} g(y_k)^T s + \frac{1}{2} s^T (H_k + \lambda_k I) s$$

but also a solution to the trust region problem

$$\min_{s \in R^n} \phi(s) = \frac{1}{2} s^T H_k s + g(y_k)^T s, \ s.t. \|s\| \le \Delta_{k,2},$$

where $\Delta_{k,2} = \left\| -(H_k + \lambda_k I)^{-1} g(y_k) \right\| = \|\widetilde{s_k}\|$.

Therefore we also have

$$\phi(0) - \phi(\widetilde{s_k}) \ge \frac{1}{2} \|g(y_k)\| \min\left\{ \|\widetilde{s_k}\|, \frac{\|g(y_k)\|}{\|H_k\|} \right\}. \quad (12)$$

Based on the inequalities (10), (11) and (12), it is reasonable for us to define the new predicted reduction as

$$\Pr ed_k = \varphi(0) - \varphi(s_k) + \phi(0) - \phi(\widetilde{s_k}), \quad (13)$$

which satisfies

$$\begin{aligned} \Pr ed_k \ge& \frac{1}{2} \|g_k\| \min\left\{ \|d_k\|, \frac{\|g_k\|}{\|H_k\|} \right\} \\ &+ \frac{1}{2} \|g(y_k)\| \min\left\{ \|\widetilde{s_k}\|, \frac{\|g(y_k)\|}{\|H_k\|} \right\}. \end{aligned} \quad (14)$$

The ratio of the actual reduction to the predicted reduction

$$r_k = \frac{Ared_k}{\Pr ed_k}, \quad (15)$$

plays a key role in deciding whether to accept the trial step and how to adjust the regularized parameter.

The modified regularized Newton algorithm with correction for unconstrained convex optimization problems is stated as follows.

Algorithm2.1

Step 1. Given $x_0 \in R^n$, $\varepsilon \ge 0$, $\mu_0 > m > 0$, $0 < p_0 \le p_1 \le p_2 < 1$. Set $k := 0$.

Step 2. If $\|g_k\| \le \varepsilon$, then stop.

Step 3. Compute $\lambda_k = \mu_k \|g_k\|$.

Solve

$$(H_k + \lambda_k I) d = -g_k \quad (16)$$

to obtain d_k.

Solve

$$(H_k + \lambda_k I) d = -g_k + \lambda_k d_k \quad (17)$$

to obtain s_k and set

$$y_k = x_k + s_k.$$

Solve

$$(H_k + \lambda_k I) s = -g(y_k) \quad (18)$$

to obtain $\widetilde{s_k}$ and set

$$t_k = s_k + \widetilde{s_k}$$

Step 4. Compute $r_k = \frac{Ared_k}{\Pr ed_k}$. Set

$$x_{k+1} = \begin{cases} x_k + t_k, & if\ r_k \ge p_0, \\ x_k, & otherwise. \end{cases} \quad (19)$$

Step 5. Choose μ_{k+1} as

$$\mu_{k+1} = \begin{cases} 4\mu_k, & if\ r_k < p_1, \\ \mu_k, & if\ r_k \in [p_1, p_2], \\ \max\{\mu_k/4, m\}, & if\ r_k > p_2. \end{cases} \quad (20)$$

Set $k := k + 1$ and go step 2.

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

Assumption 2.1 $g(x)$ and $H(x)$ are both Lipschitz continuous, that is, there exists a constant $L_1 > 0$, $L_2 > 0$ such that

$$\|g(y) - g(x)\| \le L_1 \|y - x\|, \ \forall x, y \in R^n \quad (21)$$

and

$$\|H(y) - H(x)\| \le L_2 \|y - x\|, \ \forall x, y \in R^n. \quad (22)$$

It follows from (22) that

$$\|g(y) - g(x) - H(x)(y - x)\| \le L_2 \|y - x\|^2, \ \forall x, y \in R^n. \quad (23)$$

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

Lemma2.1 A real-valued matrix A is positive semidefinite if and only if $\left(A + A^T\right)/2$ is positive semidefinite.

See [6].

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

Lemma2.2 Suppose A is positive semidefinite. Then,

$$\|A + \varphi I\| \ge \varphi$$

and

$$\left\| (A + \varphi I)^{-1} \right\| \le \varphi^{-1}$$

hold for any $\varphi > 0$.

See [14].

Theorem 2.1 Under the conditions of Assumption 2.1, if f is bounded below, then Algorithm 2.1 terminates in finite iterations or satisfies

$$\lim_{k\to\infty}\inf \|g_k\| = 0. \tag{24}$$

We prove by contradiction. If the theorem is not true, then there exists a positive τ and an integer $\widetilde{k}$ such that

$$\|g_k\| \geq \tau,\ \forall k \geq \widetilde{k}. \tag{25}$$

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

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

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 (19), we have

$$r_k < p_0,\ \forall k \geq k_1.$$

Therefore by (20) and (25), we deduce

$$\mu_k \to \infty,\ \lambda_k \to \infty, \tag{26}$$

since $x_{k+1} = x_k$, $\forall k \geq k_1$, we get from (16) and (26) that

$$\|d_k\| = \left\|-(H_k + \lambda_k I)^{-1} g_k\right\| \leq \lambda_k^{-1} \|g_k\| \to 0. \tag{27}$$

Duo to (7), we get

$$\|s_k\| = \left\|d_k + \widetilde{d_k}\right\| \leq 2\|d_k\|, \qquad \|s_k\| \to 0.$$

From (18), we obtain

$$\begin{aligned}\|\widetilde{s_k}\| &= \left\|-(H_k + \lambda_k I)^{-1} g(y_k)\right\| \\ &\leq \left\|(H_k + \lambda_k I)^{-1}(g(y_k) - g_k - H_k s_k)\right\| \\ &+ \left\|(H_k + \lambda_k I)^{-1} g_k\right\| + \left\|(H_k + \lambda_k I)^{-1} H_k s_k\right\| \\ &\leq L_2 {\lambda_k}^{-1} \|s_k\|^2 + \|d_k\| + \|s_k\| \\ &\leq \gamma_1 \|d_k\|,\end{aligned} \tag{28}$$

where γ_1 is a positive constant.

It follows from (9) and (13) that

$$\begin{aligned}&|Ared_k - \Pr ed_k| \\ &= |f(x_k) - f(x_k + s_k + \widetilde{s_k}) - (\varphi(0) - \varphi(s_k) + \phi(0) - \phi(\widetilde{s_k}))| \\ &\leq \left|f(y_k + \widetilde{s_k}) - f(y_k) - \frac{1}{2}\widetilde{s_k}^T H_k \widetilde{s_k} - g(y_k)^T \widetilde{s_k}\right| \\ &+ \left|f(y_k) - f(x_k) - \frac{1}{2}{s_k}^T H_k s_k - {g_k}^T s_k\right| \\ &\leq o\left(\|s_k\|^2\right) + o\left(\|\widetilde{s_k}\|^2\right).\end{aligned} \tag{29}$$

Moreover, from (14), (25), (21) and (27), we have

$$\Pr ed_k \geq \frac{1}{2}\tau \min\left\{\|d_k\|, \frac{\tau}{L_1}\right\} \geq \frac{1}{2}\tau \|d_k\| \tag{30}$$

for sufficiently large k.

Duo to (29) and (30), we get

$$\begin{aligned}&|r_k - 1| \\ &= \left|\frac{Ared_k - \Pr ed_k}{\Pr ed_k}\right| \\ &\leq \left|\frac{f(x_k) - f(x_k + s_k + \widetilde{s_k}) - (\varphi(0) - \varphi(s_k) + \phi(0) - \phi(\widetilde{s_k}))}{\frac{1}{2}\tau \min\left\{\|d_k\|, \frac{\tau}{L_1}\right\}}\right| \\ &\leq \frac{o\left(\|s_k\|^2\right) + o\left(\|\widetilde{s_k}\|^2\right)}{\|d_k\|} \to 0,\end{aligned} \tag{31}$$

which implies that $r_k \to 1$. Hence, there exists positive constant γ_2 such that $\mu_k \leq \gamma_2$, holds for all large k, which contradicts to (26).

Case (2): T is infinite. Then we have from (14) and (25) that

$$\begin{aligned}\infty > f(x_1) - \lim_{k\to\infty}\inf f(x_k) &\geq \sum_{i=1}^{\infty}(f(x_i) - f(x_{i+1})) \\ &= \sum_{k\in T}(f(x_k) - f(x_{k+1})) \geq \sum_{k\in T} p_0 \Pr ed_k \\ &\geq \sum_{k\in T}\frac{p_0}{2}\|g_k\| \min\left\{\|d_k\|, \frac{\|g_k\|}{\|H_k\|}\right\} \\ &+ \sum_{k\in T}\frac{p_0}{2}\|g(y_k)\| \min\left\{\|\widetilde{s_k}\|, \frac{\|g(y_k)\|}{\|H_k\|}\right\} \\ &\geq \sum_{k\in T} p_0 \frac{\tau}{2} \min\left\{\|d_k\|, \frac{\tau}{L_1}\right\},\end{aligned} \tag{32}$$

which implies that

$$\lim_{k\to\infty, k\in T} d_k = 0. \tag{33}$$

The above equality together with the updating rule of (20) means

$$\lambda_k \to \infty. \tag{34}$$

Similar to (28), it follows from (33) and (34) that

$$\|\widetilde{s_k}\| \leq \gamma_3 \|d_k\|, \|s_k\| \leq 2\|d_k\|, \quad \forall k \in T,$$

for some positive constant γ_3. Then we have

$$\|t_k\| \leq \|s_k\| + \|\widetilde{s_k}\| \leq (\gamma_3 + 2)\|d_k\|, \quad \forall k \in T.$$

This equality together with (32) yields

$$\sum_{k\in T}\|t_k\| < \infty,$$

which implies that

$$x_k \to x^*. \tag{35}$$

It follows from (16), (35), (34) and (28) that

$$s_k \to 0, \quad \widetilde{s_k} \to 0, \tag{36}$$

since $(H_k + \mu_k \|g_k\| I)\, d_k = -g_k$ from (16), we have from (21), (25) and (36) that

$$1 \leq \frac{\|H_k\|}{\|g_k\|}\|d_k\| + \mu_k \|d_k\| \leq \frac{L_1}{\tau}\|d_k\| + \mu_k \|d_k\|,$$

which means

$$\mu_k \to \infty. \tag{37}$$

By the same analysis as (31) we know that

$$r_k \to 1. \tag{38}$$

Hence, there exists a positive constant $\gamma_4 > m$ such that $\mu_k \le \gamma_4$ holds for all sufficiently large k, which gives a contradiction to (37). The proof is completed.

III. Local convergence of Algorithm 2.1

In this section, we show that the sequence generated by Algorithm 2.1 converges to some solution of (1) quadratically. To study the local convergence properties of Algorithm 2.1, we make the following assumptions.

Assumption 3.1
(a) $f(x)$ is convex and continuously differentiable.
(b) The sequence $\{x_k\}$ generated by Algorithm 2.1 converges to $x^* \in X$ and lies in some neighbourhood of x^*
(c) $\|g(x)\|$ provides a local error bound on some $N(x^*, b_1)$ for (2), that is, there exist positive constants $c_1 > 0$ and $b_1 < 1$ such that

$$\begin{aligned} &\|g(x)\| \ge c_1 dist\,(x, X) \\ &\forall x \in N(x^*, b_1) = \{x|\,\|x - x^*\| \le b_1\}. \end{aligned} \tag{39}$$

(d) The Hessian $H(x)$ is Lipschitz continuous on $N(x^*, b_1)$, i.e., there exists a positive constant $\widetilde{L_1}$ such that

$$\|H(y)-H(x)\| \le \widetilde{L_1}\,\|y-x\| \qquad \forall x, y \in N(x^*, b_1). \tag{40}$$

Note that, if $H(x)$ is nonsingular at a solution, then $\|g(x)\|$ provides a local error bound on its neighbourhood. However, the converse is not necessarily true, for examples please refer to [16] and [17] . Hence, the local error bound condition is weaker than nonsingularity.

By Assumption 3.1 (d), we know

$$\begin{aligned} &\|g(y)-g(x)-H(x)(y-x)\| \le \widetilde{L_1}\|y-x\|^2 \\ &\forall x, y \in N(x^*, b_1) \end{aligned} \tag{41}$$

and there exists a constant $\widetilde{L_2} > 0$, such that

$$\|g(y)-g(x)\| \le \widetilde{L_2}\,\|y-x\| \qquad \forall x, y \in N(x^*, b_1). \tag{42}$$

In the following, we denote $\overline{x_k}$ the vector in the solution set X that satisfies

$$\|\overline{x_k} - x_k\| = dist\,(x_k, X).$$

The following lemma gives the relationship between the trial step t_k and the distance from x_k to the solution set.

Lemma 3.1 Under the conditions of Assumption 3.1 hold. If $x_k \in N(x^*, b_1/2)$, then we have

$$\|t_k\| \le O\,(dist\,(x_k, X)). \tag{43}$$

Since $x_k \in N(x^*, b_1/2)$, we have

$$\|\overline{x_k} - x^*\| \le \|\overline{x_k} - x_k\| + \|x_k - x^*\| \le 2\,\|x_k - x^*\| \le b_1,$$

which means $\overline{x_k} \in N(x^*, b_1)$.

Then it follows from the local error bound condition yields

$$\lambda_k = \mu_k\,\|g_k\| \ge mc_1 dist(x_k, X) = mc_1\,\|\overline{x_k} - x_k\|. \tag{44}$$

From (7), we get

$$\begin{aligned} \left\|\widetilde{d_k}\right\| &= \left\|-\lambda_k (H_k + \lambda_k I)^{-1} d_k\right\| \\ &\le \lambda_k \left\|(H_k + \lambda_k I)^{-1}\right\| \|d_k\| \\ &\le \|d_k\|. \end{aligned} \tag{45}$$

Moreover, we deduce from (41), (44), Lemma2.2 and $g(\overline{x_k}) = 0$ that

$$\begin{aligned} &\|d_k - (\overline{x_k} - x_k)\| \\ &= \left\|-(H_k + \lambda_k I)^{-1} g_k - \overline{x_k} + x_k\right\| \\ &= \left\|(H_k + \lambda_k I)^{-1}(g_k + (H_k + \lambda_k I)\,(\overline{x_k} - x_k))\right\| \\ &\le \left\|(H_k + \lambda_k I)^{-1}\right\| \|g_k + H_k(\overline{x_k} - x_k)\| \\ &+\lambda_k \left\|(H_k + \lambda_k I)^{-1}\right\| \|\overline{x_k} - x_k\| \\ &\le \lambda_k^{-1} \widetilde{L_1}\|\overline{x_k} - x_k\|^2 + \|\overline{x_k} - x_k\| \\ &= O\,(\|\overline{x_k} - x_k\|) \end{aligned} \tag{46}$$

which yields

$$\|d_k\| = O\,(\|\overline{x_k} - x_k\|)\,. \tag{47}$$

Combining (45) and (47), we obtain

$$\|s_k\| = \left\|d_k + \widetilde{d_k}\right\| \le \|d_k\| + \left\|\widetilde{d_k}\right\| \le O\,(\|\overline{x_k} - x_k\|)\,, \tag{48}$$

since $y_k = x_k + s_k$, then $y_k \to x^*$, which means $y_k \in N(x^*, b_1)$ for sufficiently large k.

From (18), we get

$$\begin{aligned} \|\widetilde{s_k}\| &= \left\|-(H_k + \lambda_k I)^{-1} g\,(y_k)\right\| \\ &\le \left\|(H_k + \lambda_k I)^{-1}\,(g(y_k) - g_k - H_k s_k)\right\| \\ &+ \left\|(H_k + \lambda_k I)^{-1} g_k\right\| + \left\|(H_k + \lambda_k I)^{-1} H_k s_k\right\| \\ &\le \widetilde{L_1}\lambda_k^{-1}\|s_k\|^2 + \|d_k\| + \|s_k\| \\ &= O\,(\|\overline{x_k} - x_k\|)\,. \end{aligned} \tag{49}$$

Duo to (48) and (49), we get

$$\|t_k\| = \|s_k + \widetilde{s_k}\| \le \|s_k\| + \|\widetilde{s_k}\| \le O\,(\|\overline{x_k} - x_k\|) \tag{50}$$

The proof is completed.

Lemma 3.2 Under the conditions of Assumption 3.1, then there exists a positive constant $M > m$ such that

$$\mu_k \le M$$

holds for all sufficiently large k.

From (10), (11), (39) and (42), we have

$$\begin{aligned} &\varphi(0) - \varphi(s_k) \\ &\ge \frac{1}{2}\,\|g_k\| \min\left\{\|d_k\|\,, \frac{\|g_k\|}{\|H_k\|}\right\} \\ &\ge \frac{1}{2} c_1\,\|\overline{x_k} - x_k\| \min\left\{\|d_k\|\,, \frac{c_1\,\|\overline{x_k} - x_k\|}{\widetilde{L_2}}\right\} \\ &\ge c_2\,\|\overline{x_k} - x_k\| \min\{\|d_k\|\,, \|\overline{x_k} - x_k\|\} \end{aligned} \tag{51}$$

for some positive constant c_2.

Then from (29), (47), (48), (49) and (51), we get

$$\begin{aligned} |r_k - 1| &= \left|\frac{Ared_k - \Pr ed_k}{\Pr ed_k}\right| \\ &\le \frac{o\left(\|s_k\|^2\right) + o\left(\|\widetilde{s_k}\|^2\right)}{\|\overline{x_k} - x_k\| \min\{\|d_k\|\,, \|\overline{x_k} - x_k\|\}} \to 0, \end{aligned} \tag{52}$$

which implies that $r_k \to 1$. Therefore there exists a constant $M > m$ such that $\mu_k \le M$ holds for all sufficiently large k.

Lemma 3.3 Under the conditions of Assumption 3.1, then we have

$$dist(x_{k+1}, X) \le O(dist(x_k, X)^2)$$

From (39) and (42), we have

$$mc_1 \|\overline{x_k} - x_k\| \le \lambda_k = \mu_k \|g_k\| = \mu_k \|g_k - g(\overline{x_k})\| \\ \le M\widetilde{L_2} \|\overline{x_k} - x_k\|, \tag{53}$$

which shows that $\|\overline{x_k} - x_k\|$ is equivalent to λ_k.

From the local error bound condition, (18), (40) and (41), we have

$$\begin{aligned} & c_1 \|\overline{x_{k+1}} - x_{k+1}\| \\ \le & \|g(x_{k+1})\| = \|g(y_k + \widetilde{s_k})\| \\ \le & \|g(y_k + \widetilde{s_k}) - g(y_k) - H(y_k)\widetilde{s_k}\| \\ & + \|g(y_k) + H(y_k)\widetilde{s_k}\| \\ \le & \widetilde{L_1}\|\widetilde{s_k}\|^2 + \|g(y_k) + H_k\widetilde{s_k}\| + \|(H(y_k) - H_k)\widetilde{s_k}\| \\ \le & \widetilde{L_1}\|\widetilde{s_k}\|^2 + \lambda_k \|\widetilde{s_k}\| + \widetilde{L_1}\|s_k\|\|\widetilde{s_k}\| \\ = & O\left(\|\overline{x_k} - x_k\|^2\right) \end{aligned} \tag{54}$$

Note that since

$$\|\overline{x_k} - x_k\| \le \|\overline{x_{k+1}} - x_k\| \le \|\overline{x_{k+1}} - x_{k+1}\| + \|t_k\|$$

we may deduce from (54) that

$$\|\overline{x_k} - x_k\| \le 2\|t_k\| \tag{55}$$

for all sufficiently large k. Combining this inequality with (50) and (54), we obtain that

$$\|t_{k+1}\| = O\left(\|t_k\|^2\right), \tag{56}$$

which indicates that $\{x_k\}$ converges quadratically to x^*, namely,

$$\|x_{k+1} - x^*\| = O\left(\|x_k - x^*\|^2\right) \tag{57}$$

IV. Conclusion

In this paper, we propose a new modified regularized Newton method with correction for unconstrained convex optimization. At every iteration, not only a RNM step is computed but also two correction steps are computed which make use of the available factorization of $(H_k + \lambda_k I)$ in (16), and only need a small amount of additional calculations to obtain t_k. Under the local error bound condition, we show that the method achieves the quadratic convergence.

References

[1] Chungen Shen, Xiongda Chen, Yumei Liang, A regularized Newton method for degenerate unconstrained optimization problems, Optimization Letters, vol. 6, no. 8, pp. 1913-1933, 2012.

[2] Cyril Dennis Enyi and Mukiawa Edwin Soh, "Modified Gradient-Projection Algorithm for Solving Convex Minimization Problem in Hilbert Spaces," IAENG International Journal of Applied Mathematics, vol. 44, no. 3, pp. 144-150, 2014.

[3] Songhai Deng, ZhongWan, Xiaohong Chen, An Improved Spectral Conjugate Gradient Algorithm for Nonconvex Unconstrained Optimization Problems, Journal of Optimization Theory and Applications, vol. 157, no. 3, pp. 820-842, 2013.

[4] Yang Weiwei, Yang Yueting, Zhang Chenhui, Cao Mingyuan, A Newton-Like Trust Region Method for Large-Scale Unconstrained Nonconvex Minimization, Abstract and Applied Analysis, vol. 2013, 2013.

[5] Hao Zhang, Qin Ni, A new regularized quasi-Newton algorithm for unconstrained optimization, Applied Mathematics and Computation, 259, 2015, pp. 460-469.

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

[7] W. Zhou, D. Li, A globally convergent BFGS method for nonlinear monotone equations without any merit functions, Mathematics of Computation, vol. 77, no. 264, pp. 2231-2240, 2008.

[8] Ioaanis K, Argyros, Said Hilout, On the semilocal convergence of damped Newton's method, Applied Mathematics and Computation, vol. 219, no. 5, pp. 2808-2824, 2012.

[9] C.T.Kelley, Iterative Methods for Optimization, in: Frontiers in Applied Mathematics, vol. 18, SIAM, Philadelphia, 1999.

[10] D. Sun, A regularization Newton method for solving nonlinear complementarity problems, Applied Mathematics and Optimization, vol. 40, no. 3, pp. 315-339, 1999.

[11] D. H. Li, M. Fukushima, L. Qi, and N. Yamashita, Regularized Newton methods for convex minimization problems with singular solutions, Computational optimization and applications, vol. 28, no. 2, pp. 131-147, 2004.

[12] J.Y.Fan and Y.X.Yuan, On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption, COMPUTING, vol. 74, no. 1, pp. 23-39, 2005.

[13] J.Y.Fan, J.Y.Pan, A note on the Levenberge-Marquardt parameter, Applied Mathematics and Computation, vol. 207, no. 2, pp. 351-359, 2009.

[14] JinyanFan, YaxiangYuan, A regularized Newton method for monotone nonlinear equations and its application, Optimization Methods and Software, vol. 29, no. 1, pp. 102-119, 2014.

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

[16] N.Yamashita,M. Fukushima, On the rate of convergence of the Levenberg-Marquardt method, Computing 15 (Suppl.)(2001),pp. 227-238.

[17] Dong-huiLi, Masao Fukushima, Liqun Qi, Nobuo Yamashita, Regularized Newton Methods for Convex Minimization Problems with Singular Solutions, Computational Optimization and Applications, 28, 2004, pp. 131-147.

[18] Polyak, R.A., Regularized Newton method for unconstrained convex optimization, Mathematical Programming, 120, 2009, pp. 125-145.

[19] Weijun Zhou, Xinlong Chen, On the convergence of a modified regularized Newton method for convex optimization with singular solutions, Journal of Computational and Applied Mathematics 239, 2013, pp. 179-188.

[20] C.T.Kelley, Solving Nonlinear Equations with Newton's Method, Fundamentals of Algorithm, SIAM, Philadelphia, PA, 2003.

[21] JINYAN FAN, Accelerating the modified Levenberg-Marquardt method for nonlinear equations, Mathematics of Computation, vol. 83, no. 287, pp. 1173-1187, 2014.

[22] Y.X. Yuan, A review of trust region algorithms for optimization, in ICM99: Proceedings of the Fourth International Congress on Industrial and Applied Mathematics, J.M. Ball and J.C.R. Hunt, eds., Oxford University Press, Edinburgh, 2000, pp. 271-282.

[23] Monnanda Erappa Shobha and Santhosh George, On Improving the Semilocal Convergence of Newton-Type Iterative Method for Ill-posed Hammerstein Type Operator Equations, IAENG International Journal of Applied Mathematics, vol. 43, no. 2, pp. 64-70, 2013.

[24] Jinyan Fan, Jianyu Pan, An improved trust region algorithm for nonlinear equations, Computational Optimization and Applications, vol. 48, no. 1, pp. 59-70, 2011.