Convergence Properties of the Regularized Newton Method for the Unconstrained Nonconvex Optimization*

Kenji Ueda† and Nobuo Yamashita‡

Abstract

The regularized Newton method (RNM) is one of the efficient solution methods for the unconstrained convex optimization. It is well-known that the RNM has good convergence properties as compared to the steepest descent method and the pure Newton's method. For example, Li, Fukushima, Qi and Yamashita showed that the RNM has a quadratic rate of convergence under the local error bound condition. Recently, Polyak showed that the global complexity bound of the RNM, which is the first iteration k such that $\|\nabla f(x_k)\| \le \epsilon$, is $O(\epsilon^{-4})$, where f is the objective function and ϵ is a given positive constant. In this paper, we consider the RNM for the unconstrained "nonconvex" optimization. We show that the RNM has the following properties. (a) The RNM has a global convergence property under appropriate conditions. (b) The global complexity bound of the RNM is $O(\epsilon^{-2})$ if $\nabla^2 f$ is Lipschitz continuous on a certain compact set. (c) The RNM has a superlinear rate of convergence under the local error bound condition.

Key words. Regularized Newton methods, Global convergence, Global complexity bound, Local error bound, Superlinear convergence

1 Introduction

In this paper, we consider the regularized Newton method (RNM) for the following unconstrained minimization problem.

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\, f(x), \tag{1.1}$$

where f is a twice continuously differentiable function from $\mathbb{R}^n$ into $\mathbb{R}$. When f is convex, the RNM is one of the efficient solution methods for (1.1) and has good convergence properties [3, 4].

For a current point x_k, the RNM adopts a search direction d_k defined by

$$d_k = -(\nabla^2 f(x_k) + \mu_k I)^{-1}\nabla f(x_k), \tag{1.2}$$

where μ_k is a positive parameter. If f is convex, then $\nabla^2 f(x_k) + \mu_k I$ is a positive definite matrix, and hence d_k is a descent direction for f at x_k, i.e., $\nabla f(x_k)^T d_k < 0$. Therefore, the RNM with an appropriate line search method, such as the Armijo's step size rule, has a global convergence property.

Li, Fukushima, Qi and Yamashita [3] showed that the RNM has a quadratic rate of convergence under the assumption that $\|\nabla f(x)\|$ provides a local error bound for (1.1) in a neighborhood of an optimal solution x^*. Note that the local error bound condition holds if the second-order sufficient optimality condition holds at x^*. But the converse is not true. Thus the local error bound condition is weaker than the second-order sufficient optimality condition.

*Technical Report, 2008-015, November 26, 2008

†Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (kueda@amp.i.kyoto-u.ac.jp)

‡Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (nobuo@i.kyoto-u.ac.jp)

Recently, Polyak [4] proposed the RNM with a special step size and analyzed its global complexity bound defined as follows. Let $\{x_k\}$ be a sequence generated by some algorithms. For a given positive constant ϵ, let J be the first iteration that satisfies

$$\|\nabla f(x_J)\| \leq \epsilon.$$

We call J the global complexity bound of the algorithm. Polyak showed that if the level set of f at the initial point x_0 is compact, then the global complexity bound of his method satisfies

$$J = O(\epsilon^{-4}).$$

Note that this complexity holds without the local error bound condition or the second-order sufficient optimality condition. However, since the Polyak's RNM uses a special step size containing the Lipschitz constant of ∇f, the above result may not hold if the Lipschitz constant is unknown.

In most past studies for the RNM, the convergence properties have been discussed only when f is convex. In this paper, we consider the RNM extended to the problem (1.1) whose objective function f is nonconvex. The extended RNM (E-RNM) uses the Armijo's step size rule, and it does not contain unknown constants, e.g., the Lipschitz constant of ∇f as Polyak's method. We show that the E-RNM has the following properties:

- If a sequence $\{x_k\}$ generated by the E-RNM is bounded, then $\|\nabla f(x_k)\|$ converges to 0.
- If $\{x_k\}$ is bounded and $\nabla^2 f$ is Lipschitz continuous on a certain compact set containing $\{x_k\}$, then the global complexity bound of the E-RNM is

$$J = O(\epsilon^{-2}).$$

- Under the local error bound condition, the distance between x_k and the local optimal solution set converges to 0 superlinearly.

This paper is organized as follows. In the next section, we extend the RNM to the problem (1.1) whose objective function f is not necessarily convex. In Section 3, we show that the E-RNM has global convergence. In Section 4, we give the global complexity bound of the E-RNM. In Section 5, we establish superlinear convergence under the local error bound condition. Finally, Section 6 concludes the paper.

We use the following notations throughout the paper. For a vector $x \in \mathbb{R}^n$, $\|x\|$ denotes the Euclidean norm defined by $\|x\| := \sqrt{x^T x}$. For a symmetric matrix $M \in \mathbb{R}^{n \times n}$, we denote the maximum eigenvalue and the minimum eigenvalue of M as $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$, respectively. Then, $\|M\|$ denotes the ℓ_2 norm of M defined by $\|M\| := \sqrt{\lambda_{\max}(M^T M)}$. If M is symmetric positive semidefinite matrix, then $\|M\| = \lambda_{\max}(M)$. Furthermore, $M \succ (\succeq) 0$ denotes the positive (semi)definiteness of M, i.e., $\lambda_{\min}(M) > (\geq) 0$. $B(x, r)$ denotes the closed sphere with center x and radius r, i.e., $B(x, r) := \{y \in \mathbb{R}^n \mid \|y - x\| \leq r\}$. $\mathrm{dist}(x, S)$ denotes the distance between a vector $x \in \mathbb{R}^n$ and a set $S \subseteq \mathbb{R}^n$, i.e., $\mathrm{dist}(x, S) := \min_{y \in S} \|y - x\|$. For sets $S_1 \subseteq \mathbb{R}^n$ and $S_2 \subseteq \mathbb{R}^n$, $S_1 + S_2$ denotes the sum of S_1 and S_2 defined by $S_1 + S_2 := \{x + y \in \mathbb{R}^n \mid x \in S_1, y \in S_2\}$.

2 Extended regularized Newton method for the unconstrained nonconvex optimization

In this section, we extend the RNM to be able to used when f is nonconvex. Let x_k be the k-th iterative point. In what follows, we denote the gradient $\nabla f(x_k)$ and the Hessian $\nabla^2 f(x_k)$ as g_k and H_k, respectively. At the k-th iteration of the RNM, we set the regularized parameter μ_k as

$$\mu_k = c_1 \Lambda_k + c_2 \|g_k\|^\delta, \tag{2.1}$$

where c_1, c_2, δ are constants such that $c_1 > 1, c_2 > 0, \delta \geq 0$, and Λ_k is defined by

$$\Lambda_k := \max(0, -\lambda_{\min}(H_k)). \tag{2.2}$$

From the definition of Λ_k, the matrix $H_k + c_1\Lambda_k I$ is positive semidefinite even if f is nonconvex. Therefore, if $\|g_k\| \neq 0$, then $H_k + \mu_k I = H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I \succ 0$. So we can adopt a search direction d_k at x_k as

$$d_k = -(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1} g_k \tag{2.3}$$

when f is nonconvex.

The algorithm of the RNM with the above d_k and Armijo's step size rule is described as follows.

Extended Regularized Newton Method (E-RNM)

Step 0 : Choose parameters $\delta, c_1, c_2, \alpha, \beta$ such that

$$\delta \geq 0,\ c_1 > 1,\ c_2 > 0,\ 0 < \alpha < 1,\ 0 < \beta < 1.$$

Choose a starting point x_0. Set $k := 0$.

Step 1 : If the stopping criteria is satisfied, then terminate. Otherwise, go to Step 2.

Step 2 : [Computing search direction d_k]
Compute

$$d_k = -(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1} g_k.$$

Step 3 : [Armijo's step size rule]
Find the smallest nonnegative integer l_k such that

$$f(x_k) - f(x_k + \beta^{l_k} d_k) \geq -\alpha\beta^{l_k} g_k^T d_k. \tag{2.4}$$

Step 4 : Set $t_k = \beta^{l_k}$, $x_{k+1} = x_k + t_k d_k$ and $k := k + 1$. Go to Step 1.

In Step 3 of the E-RNM, a backtracking scheme is used. Since $d_k^T g_k < 0$ for k such that $\|g_k\| \neq 0$, the number of backtracking steps is finite.

3 Global convergence

In this section, we investigate the global convergence of the E-RNM. To this end, we need the following assumption.

Assumption 1. *There exists a compact set $\Omega \subseteq \mathbb{R}^n$ such that $\{x_k\} \subseteq \Omega$.*

Note that Assumption 1 holds if the level set of f at the initial point x_0 is compact.

First, we show the boundedness of $\{d_k\}$.

Lemma 3.1. *Suppose that $\|g_k\| \neq 0$. Then, d_k defined by (2.3) satisfies*

$$\|d_k\| \leq \frac{\|g_k\|^{1-\delta}}{c_2}.$$

Proof. We have from (2.3) that

$$\begin{aligned}
\|d_k\| &= \|(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1} g_k\| \\
&\leq \|(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1}\| \cdot \|g_k\| \\
&= \lambda_{\max}\Big((H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1}\Big)\|g_k\| \\
&= \frac{\|g_k\|}{\lambda_{\min}(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)} \\
&\leq \frac{\|g_k\|^{1-\delta}}{c_2},
\end{aligned}$$

where the last inequality follows from the facts that $H_k + c_1\Lambda_k I$ is positive semidefinite and $\|g_k\| \neq 0$. $\square$

Since $\{x_k\}$ is in the compact set Ω, there exists $U_g > 0$ such that

$$\|g_k\| \leq U_g, \quad \forall k \geq 0. \tag{3.1}$$

The next lemma indicates that $\|d_k\|$ is bounded above if $\|g_k\|$ does not converges to 0.

Lemma 3.2. *Suppose that Assumption 1 holds. Suppose also that there exists a constant $\epsilon > 0$ such that $\|g_k\| \geq \epsilon$. Then, d_k defined by (2.3) satisfies*

$$\|d_k\| \leq b(\epsilon),$$

where

$$b(\epsilon) := \max\left(\frac{U_g^{1-\delta}}{c_2}, \frac{1}{c_2\epsilon^{\delta-1}}\right).$$

Proof. When $\delta \leq 1$, it follows from Lemma 3.1 and (3.1) that

$$\|d_k\| \leq \frac{U_g^{1-\delta}}{c_2}. \tag{3.2}$$

Meanwhile, when $\delta > 1$, it follows from Lemma 3.1 and $\|g_k\| \geq \epsilon$ that

$$\|d_k\| \leq \frac{1}{c_2\epsilon^{\delta-1}}.$$

This completes the proof. $\square$

When $\|g_k\| \geq \epsilon$ for all k, we have from Lemma 3.2 that

$$x_k + \tau d_k \in \Omega + B(0, b(\epsilon)), \quad \forall \tau \in [0,1], \quad \forall k \geq 0.$$

Moreover, since $\Omega + B(0, b(\epsilon))$ is compact and f is twice continuously differentiable, there exists $U_H(\epsilon) > 0$ such that

$$\|\nabla^2 f(x)\| \leq U_H(\epsilon), \quad \forall x \in \Omega + B(0, b(\epsilon)). \tag{3.3}$$

Next, we show that the step size t_k determined in Step 4 of the E-RNM is bounded away from 0 when $\|g_k\| \geq \epsilon$.

Lemma 3.3. *Suppose that Assumption 1 holds. Suppose also that there exists a constant $\epsilon > 0$ such that $\|g_k\| \geq \epsilon$. Then,*

$$t_k \geq t_{\min}(\epsilon),$$

where

$$t_{\min}(\epsilon) := \min\left(1, \frac{2(1-\alpha)\beta c_2\epsilon^{\delta}}{U_H(\epsilon)}\right).$$

Proof. From Taylor's theorem, there exists $\tau_k \in (0,1)$ such that

$$f(x_k + t_k d_k) = f(x_k) + t_k g_k^T d_k + \frac{1}{2}t_k^2 d_k^T \nabla^2 f(x_k + \tau_k t_k d_k) d_k.$$

Thus we have

$$f(x_k) - f(x_k + t_k d_k) + \alpha t_k g_k^T d_k = (1-\alpha) t_k g_k^T d_k - \frac{1}{2}t_k^2 d_k^T \nabla^2 f(x_k + \tau_k t_k d_k) d_k. \tag{3.4}$$

Since

$$g_k = -(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)d_k$$

from the definition (2.3) of d_k, substituting it into (3.4) yields

$$\begin{aligned}
&f(x_k) - f(x_k + t_k d_k) + \alpha t_k g_k^T d_k \\
&\quad = (1-\alpha)t_k d_k^T(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)d_k - \frac{1}{2}t_k^2 d_k^T \nabla^2 f(x_k + \tau_k t_k d_k)d_k \\
&\quad = (1-\alpha)t_k d_k^T\left(H_k + c_1\Lambda_k I\right)d_k + (1-\alpha)t_k d_k^T\left(c_2\|g_k\|^\delta I - \frac{1}{2(1-\alpha)}t_k\nabla^2 f(x_k + \tau_k t_k d_k)\right)d_k. \quad (3.5)
\end{aligned}$$

Since $H_k + c_1\Lambda_k I$ is positive semidefinite, we have

$$\begin{aligned}
f(x_k) - f(x_k + t_k d_k) + \alpha t_k g_k^T d_k &\geq (1-\alpha)t_k d_k^T\left(c_2\|g_k\|^\delta I - \frac{1}{2(1-\alpha)}t_k\nabla^2 f(x_k + \tau_k t_k d_k)\right)d_k \\
&\geq (1-\alpha)t_k\left(c_2\|g_k\|^\delta - \frac{1}{2(1-\alpha)}t_k\|\nabla^2 f(x_k + \tau_k t_k d_k)\|\right)\|d_k\|^2. \quad (3.6)
\end{aligned}$$

It then follows from $\|g_k\| \geq \epsilon$ and (3.3) that

$$f(x_k) - f(x_k + t_k d_k) + \alpha t_k g_k^T d_k \geq (1-\alpha)t_k\left(c_2\epsilon^\delta - \frac{1}{2(1-\alpha)}t_k U_H(\epsilon)\right)\|d_k\|^2. \quad (3.7)$$

Now we consider two cases: (i) $2(1-\alpha)c_2\epsilon^\delta/U_H(\epsilon) \geq 1$ and (ii) $2(1-\alpha)c_2\epsilon^\delta/U_H(\epsilon) < 1$.

Case (i): From (3.7), we have

$$f(x_k) - f(x_k + d_k) \geq -\alpha g_k^T d_k,$$

and hence $t_k = 1$ satisfies the Armijo's rule (2.4).

Case (ii): In this case, it follows from (3.7) that

$$t_k \leq \frac{2(1-\alpha)c_2\epsilon^\delta}{U_H(\epsilon)} \Rightarrow f(x_k) - f(x_k + t_k d_k) \geq -\alpha t_k g_k^T d_k.$$

Therefore, t_k must be

$$t_k \geq \left(\frac{2(1-\alpha)c_2\epsilon^\delta}{U_H(\epsilon)}\right)\beta.$$

Otherwise t_k/β satisfies the Armijo's rule (2.4), which contradicts the definition of t_k.

This completes the proof. □

Next, we give a lower bound of the reduction $f(x_k) - f(x_{k+1})$ when $\|g_k\| \geq \epsilon$.

Lemma 3.4. *Suppose that Assumption 1 holds. Suppose also that there exists a constant $\epsilon > 0$ such that $\|g_k\| \geq \epsilon$. Then,*

$$f(x_k) - f(x_{k+1}) \geq p(\epsilon)\epsilon^2,$$

where

$$p(\epsilon) := \frac{\alpha t_{\min}(\epsilon)}{(1+c_1)U_H(\epsilon) + c_2 U_g^\delta}.$$

Proof. Since $H_k + c_1\Lambda_k I$ is positive semidefinite and $\|g_k\| \neq 0$, we have

$$\begin{aligned}
\lambda_{\min}\Big((H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1}\Big) &= \frac{1}{\lambda_{\max}(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)} \\
&= \frac{1}{\lambda_{\max}(H_k) + c_1\Lambda_k + c_2\|g_k\|^\delta}.
\end{aligned}$$

It then follows from $\|g_k\| \geq \epsilon$, (3.1) and (3.3) that

$$\lambda_{\min}\Big((H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1}\Big) \geq \frac{1}{(1+c_1)U_H(\epsilon) + c_2 U_g^\delta}. \tag{3.8}$$

Therefore, we have from the Armijo's rule (2.4) and the definition (2.3) of d_k that

$$\begin{aligned}
f(x_k) - f(x_{k+1}) &\geq -\alpha t_k g_k^T d_k \\
&= \alpha t_k g_k^T (H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1} g_k \\
&\geq \alpha t_k \lambda_{\min}\Big((H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1}\Big)\|g_k\|^2 \\
&\geq \frac{\alpha t_{\min}(\epsilon)}{(1+c_1)U_H(\epsilon) + c_2 U_g^\delta}\|g_k\|^2 \qquad (3.9) \\
&\geq \frac{\alpha t_{\min}(\epsilon)}{(1+c_1)U_H(\epsilon) + c_2 U_g^\delta}\epsilon^2,
\end{aligned}$$

where the third inequality follows from (3.8) and Lemma 3.3, and the last inequality follows from $\|g_k\| \geq \epsilon$. □

From the above lemma, we show the global convergence of the E-RNM.

Theorem 3.1. *Suppose that Assumption 1 holds. Then,*

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

Proof. Suppose the contrary, i.e., $\limsup_{k\to\infty} \|g_k\| > 0$. Let

$$\begin{aligned}
\epsilon &:= \frac{\limsup_{k\to\infty}\|g_k\|}{2}, \\
I_\epsilon(k) &:= \{j \in \{0,1,\dots\} \mid j \leq k,\ \|g_j\| \geq \epsilon\}.
\end{aligned}$$

Then, we have

$$\lim_{k\to\infty} |I_\epsilon(k)| = \infty,$$

where $|I_\epsilon(k)|$ denotes the number of the elements of $I_\epsilon(k)$. From Lemma 3.4, we obtain

$$\begin{aligned}
f(x_0) - f(x_{k+1}) &\geq \sum_{j=0}^{k}(f(x_j) - f(x_{j+1})) \\
&\geq \sum_{j\in I_\epsilon(k)} (f(x_j) - f(x_{j+1})) \\
&\geq \sum_{j\in I_\epsilon(k)} p(\epsilon)\epsilon^2 \\
&= p(\epsilon)\epsilon^2\,|I_\epsilon(k)|\,.
\end{aligned}$$

Taking $k \to \infty$, the right hand side of the inequality goes to infinity. This contradicts Assumption 1 and the continuity of f. Hence, we have $\limsup_{k\to\infty}\|g_k\| = 0$, i.e., $\lim_{k\to\infty}\|g_k\| = 0$. □

4 Global complexity bound

In this section, we estimate the global complexity bound of the E-RNM. To this end, we need the following assumptions in addition to Assumption 1 in the previous section.

Assumption 2.

(a) $\delta \leq 1/2$.

(b) $\alpha \leq 1/2$.

(c) *Let* $b_1 := U_g^{1-\delta}/c_1$. $\nabla^2 f$ *is Lipschitz continuous on* $\Omega + B(0, b_1)$*, i.e., there exists* $L_H > 0$ *such that*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L_H \|x - y\|, \quad \forall x, y \in \Omega + B(0, b_1).$$

Under Assumption 1, the inequality (3.1) holds. Moreover, there exists $f_{\min}$ such that

$$f(x_k) \geq f_{\min}, \quad \forall k \geq 0. \tag{4.1}$$

From Assumptions 1 and 2 (a), the inequality (3.2) holds. Therefore, we have

$$x_k + \tau d_k \in \Omega + B(0, b_1), \quad \forall \tau \in [0, 1], \quad \forall k \geq 0.$$

Since $\Omega + B(0, b_1)$ is compact and f is twice continuously differentiable, there exists $U_H > 0$ such that

$$\|\nabla^2 f(x)\| \leq U_H, \quad \forall x \in \Omega + B(0, b_1). \tag{4.2}$$

Moreover, from Assumption 2 (c), we have

$$\|\nabla^2 f(y)(x - y) - (\nabla f(x) - \nabla f(y))\| \leq \frac{1}{2} L_H \|x - y\|^2, \quad \forall x, y \in \Omega + B(0, b_1). \tag{4.3}$$

The next lemma indicates that the step size t_k is bounded below by some positive constant independent of k.

Lemma 4.1. *Suppose that Assumptions 1 and 2 hold. Then,*

$$t_k \geq t_{\min},$$

where

$$t_{\min} := \min \left(1, \sqrt{\frac{2(1-\alpha)\beta c_2^2}{L_H U_g^{1-2\delta}}}\right).$$

Proof. Since $H_k + c_1 \Lambda_k I$ is positive semidefinite and $1 \geq \frac{1}{2(1-\alpha)} t_k$, we have

$$\begin{aligned} d_k^T (H_k + c_1 \Lambda_k) d_k &\geq \frac{1}{2(1-\alpha)} t_k d_k^T (H_k + c_1 \Lambda_k) d_k \\ &\geq \frac{1}{2(1-\alpha)} t_k d_k^T H_k d_k. \end{aligned} \tag{4.4}$$

It then follows from (3.5) and (4.4) that

$$\begin{aligned} & f(x_k) - f(x_k + t_k d_k) + \alpha t_k g_k^T d_k \\ & \quad \geq \frac{1}{2} t_k^2 d_k^T H_k d_k + (1-\alpha) t_k d_k^T \left(c_2 \|g_k\|^\delta I - \frac{1}{2(1-\alpha)} t_k \nabla^2 f(x_k + \tau_k t_k d_k) \right) d_k \\ & \quad \geq (1-\alpha) t_k \left(c_2 \|g_k\|^\delta - \frac{1}{2(1-\alpha)} t_k \|\nabla^2 f(x_k + \tau_k t_k d_k) - H_k\| \right) \|d_k\|^2. \end{aligned}$$

Moreover, since $\tau_k \in (0,1)$ and $\nabla^2 f$ is Lipschitz continuous from Assumption 2 (c), we have

$$
\begin{aligned}
f(x_k) - f(x_k + t_k d_k) + \alpha t_k g_k^T d_k &\geq (1-\alpha) t_k \left(c_2 \|g_k\|^\delta - \frac{L_H}{2(1-\alpha)} t_k^2 \|d_k\| \right) \|d_k\|^2 \\
&= \frac{L_H t_k}{2} \left(\frac{2(1-\alpha) c_2 \|g_k\|^\delta}{L_H \|d_k\|} - t_k^2 \right) \|d_k\|^3. \qquad (4.5)
\end{aligned}
$$

From Assumption 2 (a), Lemma 3.1 and (3.1), we have

$$
\frac{\|d_k\|}{\|g_k\|^\delta} \leq \frac{\|g_k\|^{1-2\delta}}{c_2} \leq \frac{U_g^{1-2\delta}}{c_2}. \qquad (4.6)
$$

Thus we obtain from (4.5) and (4.6) that

$$
f(x_k) - f(x_k + t_k d_k) + \alpha t_k g_k^T d_k \geq \frac{L_H t_k}{2} \left(\frac{2(1-\alpha) c_2^2}{L_H U_g^{1-2\delta}} - t_k^2 \right) \|d_k\|^3.
$$

If $2(1-\alpha)c_2^2/(L_H U_g^{1-2\delta}) \geq 1$, then Armijo's rule (2.4) holds with $t_k = 1$. If $2(1-\alpha)c_2^2/(L_H U_g^{1-2\delta}) < 1$, then we have

$$
t_k \leq \sqrt{\frac{2(1-\alpha) c_2^2}{L_H U_g^{1-2\delta}}} \Rightarrow f(x_k) - f(x_k + t_k d_k) \geq -\alpha t_k g_k^T d_k.
$$

Thus t_k must satisfy

$$
t_k \geq \left(\sqrt{\frac{2(1-\alpha) c_2^2}{L_H U_g^{1-2\delta}}} \right) \beta.
$$

This completes the proof. □

From the above lemma, we show that the number of backtracking steps is bounded above by some positive constant independent of k.

Theorem 4.1. *Suppose that Assumptions 1 and 2 hold. Then,*

$$
l_k \leq l_{\max},
$$

where

$$
l_{\max} := \frac{\ln t_{\min}}{\ln \beta}.
$$

Proof. From Lemma 4.1, we obtain $l_k \ln \beta \geq \ln t_{\min}$. Since $\ln \beta < 0$, we have $l_k \leq \frac{\ln t_{\min}}{\ln \beta}$. □

Remark 4.1. *Since the Polyak's RNM [4] uses a special step size which contains the Lipschitz constant, his method does not need a backtracking scheme. However, the step size used in the Polyak's RNM cannot be used when the Lipschitz constant is unknown.*

Next, we estimate a lower bound of the reduction $f(x_k) - f(x_{k+1})$.

Lemma 4.2. *Suppose that Assumptions 1 and 2 hold. Then,*

$$
f(x_k) - f(x_{k+1}) \geq p \|g_k\|^2,
$$

where

$$
p := \frac{\alpha t_{\min}}{(1+c_1) U_H + c_2 U_g^\delta}.
$$

Proof. It directly follows from (4.2), Lemma 4.1 and the inequality (3.9) of Lemma 3.4. □

By Lemma 4.2, we obtain the global complexity bound of the E-RNM.

Theorem 4.2. *Suppose that Assumptions 1 and 2 hold. Let $\{x_k\}$ be a sequence generated by the E-RNM. Let J be the first iteration such that $\|g_J\| \le \epsilon$. Then,*

$$J \le \frac{f(x_0) - f_{\min}}{p}\epsilon^{-2},$$

where p is a constant given in Lemma 4.2.

Proof. It follows from Lemma 4.2 that

$$f(x_0) - f_{\min} \ge f(x_0) - f(x_k) \ge \sum_{j=0}^{k-1}(f(x_j) - f(x_{j+1})) \ge p\sum_{j=0}^{k-1}\|g_j\|^2 \ge kp\left(\min_{0\le j\le k-1}\|g_j\|\right)^2.$$

Then, we have

$$\min_{0\le j\le k-1}\|g_j\| \le \left(\frac{f(x_0) - f_{\min}}{kp}\right)^{\frac{1}{2}},$$

and hence

$$k \ge \frac{f(x_0) - f_{\min}}{p}\epsilon^{-2} \Rightarrow \min_{0\le j\le k-1}\|g_j\| \le \epsilon.$$

This completes the proof. □

Remark 4.2. *The global complexity bound $O(\epsilon^{-2})$ given in Theorem 4.2 is better than the existing result $O(\epsilon^{-4})$ by Polyak [4]. Note that [4] does not assume Assumption 2 (c), i.e., the Lipschitz continuity of $\nabla^2 f$.*

5 Local Convergence

In this section, we show that the E-RNM has a superlinear convergence under the local error bound condition. In order to prove the superlinear convergence, we use techniques similar to [5] and [1]. In [5], Yamashita and Fukushima showed that the Levenberg-Marquardt method has a quadratic rate of convergence under the local error bound condition. Similarly, in [1], Dan, Yamashita and Fukushima showed that the inexact Levenberg-Marquardt method has a superlinear rate of convergence under the local error bound condition.

First, we make the following assumptions.

Assumption 3.

(a) *There exists a local optimal solution x^* of the problem (1.1).*

(b) $\nabla^2 f$ *is local Lipschitz continuous, i.e., there exist constants $b_2 \in (0, 1)$ and $\bar{L}_H > 0$ such that*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \le \bar{L}_H\|x - y\|, \quad \forall x, y \in B(x^*, b_2).$$

(c) $\|\nabla f(x)\|$ *provides a local error bound for the problem (1.1) on $B(x^*, b_2)$, i.e., there exists a constant $\kappa_1 > 0$ such that*

$$\kappa_1 \mathrm{dist}(x, X^*) \le \|\nabla f(x)\|, \quad \forall x \in B(x^*, b_2),$$

where X^* is the local optimal solution set of (1.1).

(d) $0 < \delta < 1$.

(e) $\alpha \le 1/2$.

Note that under Assumption 3 (b), the following inequality holds.

$$\|\nabla^2 f(y)(x-y) - (\nabla f(x) - \nabla f(y))\| \le \frac{1}{2}\bar{L}_H \|x-y\|^2, \quad \forall x, y \in B(x^*, b_2). \tag{5.1}$$

Moreover, since f is twice continuously differentiable, there exist positive constants $\bar{U}_g$ and $\bar{L}_g$ such that

$$\|\nabla f(x)\| \le \bar{U}_g, \quad \forall x \in B(x^*, b_2), \tag{5.2}$$

$$\|\nabla f(x) - \nabla f(y)\| \le \bar{L}_g \|x-y\|, \quad \forall x, y \in B(x^*, b_2). \tag{5.3}$$

In what follows, $\bar{x}_k$ denotes an arbitrary vector such that

$$\|x_k - \bar{x}_k\| = \operatorname{dist}(x_k, X^*), \quad \bar{x}_k \in X^*.$$

In the case where f is convex, Li, Fukushima, Qi and Yamashita [3] showed the RNM has a quadratic rate of convergence under the local error bound condition. The convexity of f implies $\Lambda_k \equiv 0$. However, since f is not necessarily convex, it is not always true that $\Lambda_k = 0$. Therefore, we now investigate the relationship between Λ_k and $\operatorname{dist}(x_k, X^*)$. To this end, we need the following property on a singular matrix.

Lemma 5.1. *Suppose that $M \in \mathbb{R}^{n\times n}$ is singular, then $\|I - M\| \ge 1$.*

Proof. It directly follows from [2, Corollary 5.6.16]. □

By using Lemma 5.1, we show the following key lemma for superlinear convergence.

Lemma 5.2. *Suppose that Assumption 3 holds. If $x_k \in B(x^*, b_2/2)$, then*

$$\Lambda_k \le \bar{L}_H \operatorname{dist}(x_k, X^*).$$

Proof. When $H_k \succeq 0$, we have $\Lambda_k = 0$. Thus the desired inequality holds. Next, we assume $\lambda_{\min}(H_k) < 0$. Let $\bar{\lambda}_k^{(l)}$ be the l-th largest eigenvalue of $\nabla^2 f(\bar{x}_k)$. Since $\bar{x}_k \in X^*$, we have $\bar{\lambda}_k^{(l)} \ge 0$. Moreover, since $\nabla^2 f(\bar{x}_k)$ is a real symmetric matrix, $\nabla^2 f(\bar{x}_k)$ can be diagonalized by some orthogonal matrix $\bar{Q}_k$, i.e.,

$$\bar{Q}_k^T \nabla^2 f(\bar{x}_k) \bar{Q}_k = \operatorname{diag}(\bar{\lambda}_k^{(l)}),$$

where $\operatorname{diag}(\bar{\lambda}_k^{(l)})$ denotes the diagonal matrix whose (l, l) element is $\bar{\lambda}_k^{(l)}$. Then, we obtain

$$\begin{aligned}
\lambda_{\min}(H_k)I - \bar{Q}_k^T H_k \bar{Q}_k &= \lambda_{\min}(H_k)I - \bar{Q}_k^T\Big(\nabla^2 f(\bar{x}_k) + (H_k - \nabla^2 f(\bar{x}_k))\Big)\bar{Q}_k \\
&= \lambda_{\min}(H_k)I - \operatorname{diag}(\bar{\lambda}_k^{(l)}) - \bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k.
\end{aligned}$$

Since $\bar{Q}_k^T H_k \bar{Q}_k$ has the eigenvalue $\lambda_{\min}(H_k)$, $\lambda_{\min}(H_k)I - \bar{Q}_k^T H_k \bar{Q}_k$ is singular. Thus $\lambda_{\min}(H_k)I - \operatorname{diag}(\bar{\lambda}_k^{(l)}) - \bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k$ is also singular. On the other hand, $\lambda_{\min}(H_k)I - \operatorname{diag}(\bar{\lambda}_k^{(l)})$ is nonsingular, because $\lambda_{\min}(H_k) < 0$ and $\bar{\lambda}_k^{(l)} \ge 0$.

Now let

$$M := \Big(\lambda_{\min}(H_k)I - \operatorname{diag}(\bar{\lambda}_k^{(l)})\Big)^{-1}\Big(\lambda_{\min}(H_k)I - \operatorname{diag}(\bar{\lambda}_k^{(l)}) - \bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k\Big).$$

Then, M is singular. It then follows from Lemma 5.1 that

$$\begin{aligned}
1 &\le \|I - M\| \\
&= \left\| I - \left(I - \Big(\lambda_{\min}(H_k)I - \operatorname{diag}(\bar{\lambda}_k^{(l)})\Big)^{-1} \bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k \right) \right\| \\
&= \left\| \Big(\lambda_{\min}(H_k)I - \operatorname{diag}(\bar{\lambda}_k^{(l)})\Big)^{-1} \bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k \right\| \\
&\le \left\| \Big(\lambda_{\min}(H_k)I - \operatorname{diag}(\bar{\lambda}_k^{(l)})\Big)^{-1} \right\| \cdot \|\bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k\| \\
&= \left\| \Big(\lambda_{\min}(H_k)I - \operatorname{diag}(\bar{\lambda}_k^{(l)})\Big)^{-1} \right\| \cdot \|H_k - \nabla^2 f(\bar{x}_k)\|.
\end{aligned} \tag{5.4}$$

We consider $\|(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)}))^{-1}\|$ and $\|H_k - \nabla^2 f(\bar{x}_k)\|$ separately. Since $\lambda_{\min}(H_k) < 0$ and $\bar{\lambda}_k^{(l)} \geq 0$, we have

$$\begin{aligned}
\left\| \left(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(l)}) \right)^{-1} \right\| &= \max_{1 \leq l \leq n} \left| \lambda_{\min}(H_k) - \bar{\lambda}_k^{(l)} \right|^{-1} \\
&= \frac{1}{\min_{1 \leq l \leq n} \left| \lambda_{\min}(H_k) - \bar{\lambda}_k^{(l)} \right|} \\
&\leq \frac{1}{|\lambda_{\min}(H_k)|} \\
&= \frac{1}{\Lambda_k}.
\end{aligned} \tag{5.5}$$

Next, we consider $\|H_k - \nabla^2 f(\bar{x}_k)\|$. Since $x_k \in B(x^*, b_2/2)$, we have

$$\|\bar{x}_k - x^*\| \leq \|\bar{x}_k - x_k\| + \|x_k - x^*\| \leq \|x^* - x_k\| + \|x_k - x^*\| \leq b_2,$$

and hence $\bar{x}_k \in B(x^*, b_2)$. It then follows from Assumption 3 (b) that

$$\|H_k - \nabla^2 f(\bar{x}_k)\| \leq \bar{L}_H \|x_k - \bar{x}_k\| = \bar{L}_H \mathrm{dist}(x_k, X^*). \tag{5.6}$$

Therefore, we have from (5.4) – (5.6) that

$$1 \leq \frac{\bar{L}_H \mathrm{dist}(x_k, X^*)}{\Lambda_k},$$

which is the desired inequality. □

Using this lemma, we can show the superlinear convergence in a way similar to [5] and [1]. For the completeness, we give the proofs.

Lemma 5.3. *Suppose that Assumption 3 holds. If $x_k \in B(x^*, b_2/2)$, then*

$$\|d_k\| \leq \kappa_2 \mathrm{dist}(x_k, X^*),$$

where

$$\kappa_2 := \frac{\bar{L}_H}{2 c_2 \kappa_1^\delta} + \max\left(1, \frac{1}{c_1 - 1}\right).$$

Proof. First note that $\nabla f(\bar{x}_k) = 0$. From the definition (1.2) of d_k we have

$$\begin{aligned}
\|d_k\| &= \left\|(H_k + c_1 \Lambda_k I + c_2 \|g_k\|^\delta I)^{-1} g_k\right\| \\
&= \left\|(H_k + c_1 \Lambda_k I + c_2 \|g_k\|^\delta I)^{-1} \left(g_k - \nabla f(\bar{x}_k) - H_k(x_k - \bar{x}_k) + H_k(x_k - \bar{x}_k) \right)\right\| \\
&\leq \left\|(H_k + c_1 \Lambda_k I + c_2 \|g_k\|^\delta I)^{-1} \left(g_k - \nabla f(\bar{x}_k) - H_k(x_k - \bar{x}_k) \right)\right\| \\
&\quad + \left\|(H_k + c_1 \Lambda_k I + c_2 \|g_k\|^\delta I)^{-1} H_k(x_k - \bar{x}_k)\right\| \\
&\leq \left\|(H_k + c_1 \Lambda_k I + c_2 \|g_k\|^\delta I)^{-1}\right\| \cdot \|g_k - \nabla f(\bar{x}_k) - H_k(x_k - \bar{x}_k)\| \\
&\quad + \left\|(H_k + c_1 \Lambda_k I + c_2 \|g_k\|^\delta I)^{-1} H_k\right\| \cdot \|x_k - \bar{x}_k\| \\
&\leq \frac{\bar{L}_H}{2} \|x_k - \bar{x}_k\|^2 \cdot \left\|(H_k + c_1 \Lambda_k I + c_2 \|g_k\|^\delta I)^{-1}\right\| + \|x_k - \bar{x}_k\| \cdot \left\|(H_k + c_1 \Lambda_k I + c_2 \|g_k\|^\delta I)^{-1} H_k\right\| \\
&= \frac{\bar{L}_H}{2} \mathrm{dist}(x_k, X^*)^2 \left\|(H_k + c_1 \Lambda_k I + c_2 \|g_k\|^\delta I)^{-1}\right\| + \mathrm{dist}(x_k, X^*) \left\|(H_k + c_1 \Lambda_k I + c_2 \|g_k\|^\delta I)^{-1} H_k\right\|,
\end{aligned} \tag{5.7}$$

where the last inequality follows from (5.1). First, we consider $\|(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1}\|$. Since $x_k \in B(x^*, b_2/2)$, we have $\bar{x}_k \in B(x^*, b_2)$. It follows from $H_k + c_2\Lambda_k \succeq 0$ and Assumption 3 (c) that

$$\begin{aligned}\left\|(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1}\right\| &= \lambda_{\max}\Big((H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1}\Big)\\ &= \frac{1}{\lambda_{\min}(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)}\\ &\le \frac{1}{c_2\|g_k\|^\delta}\\ &= \frac{1}{c_2\kappa_1^\delta \mathrm{dist}(x_k, X^*)^\delta}.\end{aligned} \tag{5.8}$$

Next, we consider $\|(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1}H_k\|$. Let $\lambda_k^{(l)}$ be the l-th largest eigenvalue of H_k. Then, the eigenvalues of $(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1}H_k$ are given by

$$\frac{\lambda_k^{(l)}}{\lambda_k^{(l)} + c_1\Lambda_k + c_2\|g_k\|^\delta}, \quad 1 \le l \le k.$$

Now we consider two cases: (i) $\lambda_k^{(l)} \ge 0$ and (ii) $\lambda_k^{(l)} < 0$.

Case (i): This case implies that

$$\frac{\left|\lambda_k^{(l)}\right|}{\left|\lambda_k^{(l)} + c_1\Lambda_k + c_2\|g_k\|^\delta\right|} \le 1.$$

Case (ii): In this case, since $-\Lambda_k = \lambda_{\min}(H_k) \le \lambda_k^{(l)} < 0$, we have $\lambda_k^{(l)} - \lambda_{\min}(H_k) \ge 0$ and $|\lambda_k^{(l)}| \le |\lambda_{\min}(H_k)|$. Therefore, we have

$$\begin{aligned}\frac{\left|\lambda_k^{(l)}\right|}{\left|\lambda_k^{(l)} + c_1\Lambda_k + c_2\|g_k\|^\delta\right|} &= \frac{\left|\lambda_k^{(l)}\right|}{\left|(\lambda_k^{(l)} - \lambda_{\min}(H_k)) - (c_1 - 1)\lambda_{\min}(H_k) + c_2\|g_k\|^\delta\right|}\\ &\le \frac{|\lambda_{\min}(H_k)|}{\lambda_k^{(l)} - \lambda_{\min}(H_k) + (c_1 - 1)\,|\lambda_{\min}(H_k)| + c_2\|g_k\|^\delta}\\ &\le \frac{1}{c_1 - 1}.\end{aligned}$$

Thus we have

$$\frac{\left|\lambda_k^{(l)}\right|}{\left|\lambda_k^{(l)} + c_1\Lambda_k + c_2\|g_k\|^\delta\right|} \le \max\left(1, \frac{1}{c_1 - 1}\right), \quad 1 \le l \le k,$$

and hence

$$\left\|(H_k + c_1\Lambda_k I + c_2\|g_k\|^\delta I)^{-1}H_k\right\| \le \max\left(1, \frac{1}{c_1 - 1}\right). \tag{5.9}$$

From (5.7) – (5.9), we have

$$\begin{aligned}\|d_k\| &\le \frac{\bar{L}_H}{2c_2\kappa_1^\delta}\mathrm{dist}(x_k, X^*)^{2-\delta} + \max\left(1, \frac{1}{c_1 - 1}\right)\mathrm{dist}(x_k, X^*)\\ &\le \left(\frac{\bar{L}_H}{2c_2\kappa_1^\delta} + \max\left(1, \frac{1}{c_1 - 1}\right)\right)\mathrm{dist}(x_k, X^*),\end{aligned}$$

which is the desired inequality. $\square$

From the above lemma, we can show that $x_k + \tau d_k \in B(x^*, b_2)$ for any $\tau \in [0, 1]$ if x_k is sufficiently close to x^*.

Lemma 5.4. *Suppose that Assumption 3 holds. Let $b_3 := b_2/(\kappa_2 + 1)$. If $x_k \in B(x^*, b_3)$, then*

$$x_k + \tau d_k \in B(x^*, b_2), \quad \forall \tau \in [0, 1].$$

Proof. Since $b_3 \leq b_2/2$, we have $x_k \in B(x^*, b_2/2)$. Therefore, we obtain

$$\begin{aligned} \|x_k + \tau d_k - x^*\| &\leq \|x_k - x^*\| + \|d_k\| \\ &\leq \|x_k - x^*\| + \kappa_2 \mathrm{dist}(x_k, X^*) \\ &\leq \|x_k - x^*\| + \kappa_2 \|x_k - x^*\| \\ &\leq (\kappa_2 + 1) b_3 \leq b_2, \end{aligned}$$

where the second inequality follows from Lemma 5.3. □

From the above lemma, $l_k = 0$ (that is, $t_k = 1$) is accepted in Step 3 of the RNM if x_k is sufficiently close to x^*.

Lemma 5.5. *Suppose that Assumption 3 holds. Let*

$$b_4 := \min\left(b_3, \left(\frac{2(1-\alpha)c_2\kappa_1^\delta}{\kappa_2 \bar{L}_H}\right)^{\frac{1}{1-\delta}}\right).$$

If $x_k \in B(x^*, b_4)$, then $t_k = 1$.

Proof. From Assumption 3 (b) and the inequality (4.4) we have

$$f(x_k) - f(x_k + t_k d_k) + \alpha t_k g_k^T d_k \geq \frac{\bar{L}_H t_k}{2}\left(\frac{2(1-\alpha)c_2\|g_k\|^\delta}{\bar{L}_H\|d_k\|} - t_k^2\right)\|d_k\|^3.$$

It then follows from Assumption 3 (c) and Lemma 5.3 that

$$\begin{aligned} f(x_k) - f(x_k + t_k d_k) + \alpha t_k g_k^T d_k &\geq \frac{\bar{L}_H t_k}{2}\left(\frac{2(1-\alpha)c_2\kappa_1^\delta}{\kappa_2 \bar{L}_H \mathrm{dist}(x_k, X^*)^{1-\delta}} - t_k^2\right)\|d_k\|^3 \\ &\geq \frac{\bar{L}_H t_k}{2}\left(\frac{2(1-\alpha)c_2\kappa_1^\delta}{\kappa_2 \bar{L}_H \|x_k - x^*\|^{1-\delta}} - t_k^2\right)\|d_k\|^3 \\ &\geq \frac{\bar{L}_H t_k}{2}(1 - t_k^2)\|d_k\|^3, \end{aligned}$$

where the last inequality follows from $x_k \in B(x^*, b_4)$. Therefore, we have $t_k = 1$. □

Next, we show that $\mathrm{dist}(x_k, X^*)$ converges to 0 superlinearly, as long as $\{x_k\}$ lie in a neighborhood of x^*.

Lemma 5.6. *Suppose that Assumption 3 holds. If $x_k, x_{k+1} \in B(x^*, b_4)$, then*

$$\mathrm{dist}(x_{k+1}, X^*) = O\left(\mathrm{dist}(x_k, X^*)^{1+\delta}\right).$$

Therefore, there exists a positive constant b_5 such that

$$\mathrm{dist}(x_k, X^*) \leq b_5 \Rightarrow \mathrm{dist}(x_{k+1}, X^*) \leq \frac{1}{2}\mathrm{dist}(x_k, X^*).$$

Proof. We have from Assumption 3 (c)

$$\begin{aligned}
\mathrm{dist}(x_{k+1}, X^*) &\le \frac{1}{\kappa_1}\|g_{k+1}\| \\
&\le \frac{1}{\kappa_1}\|H_k d_k + g_k\| + \frac{\bar{L}_H}{2\kappa_1}\|d_k\|^2 \\
&= \frac{1}{\kappa_1}\left\|c_1\Lambda_k d_k + c_2\|g_k\|^\delta d_k\right\| + \frac{\bar{L}_H}{2\kappa_1}\|d_k\|^2 \\
&\le \frac{c_1\Lambda_k}{\kappa_1}\|d_k^*\| + \frac{c_2}{\kappa_1}\|g_k\|^\delta\|d_k\| + \frac{\bar{L}_H}{2\kappa_1}\|d_k\|^2, \qquad (5.10)
\end{aligned}$$

where the second inequality follows from (5.1) and Lemma 5.5, the first equality follows from the definition (2.3) of d_k. From (5.3), we have

$$\|g_k\|^\delta = \|g_k - \nabla f(\bar{x}_k)\|^\delta \le \bar{L}_g^\delta \mathrm{dist}(x_k, X^*)^\delta. \qquad (5.11)$$

Therefore, we obtain from (5.10), (5.11), Lemma 5.2 and Lemma 5.3 that

$$\begin{aligned}
\mathrm{dist}(x_{k+1}, X^*) &\le \frac{c_1\kappa_2\bar{L}_H}{\kappa_1}\mathrm{dist}(x_k, X^*)^2 + \frac{c_2\kappa_2\bar{L}_g^\delta}{\kappa_1}\mathrm{dist}(x_k, X^*)^{1+\delta} + \frac{\kappa_2^2\bar{L}_H}{2\kappa_1}\mathrm{dist}(x_k, X^*)^2 \\
&\le \frac{\kappa_2(2c_1\bar{L}_H + 2c_2\bar{L}_g^\delta + \kappa_2\bar{L}_H)}{2\kappa_1}\mathrm{dist}(x_k, X^*)^{1+\delta}.
\end{aligned}$$

□

Lemma 5.6 shows that $\{\mathrm{dist}(x_k, X^*)\}$ converges to 0 superlinearly if $x_k \in B(x^*, b_4)$ for all k. Now we give a sufficient condition for $x_k \in B(x^*, b_4)$ for all k.

Lemma 5.7. *Suppose that Assumption 3 holds. Let $b_6; = \min(b_4, b_5)$ and $b_7 := \frac{1}{1+2\kappa_2}b_6$. If $x_0 \in B(x^*, b_7)$, then $x_k \in B(x^*, b_6)$ for all k.*

Proof. We prove the lemma by induction. First we consider the case $k = 0$. Since $b_7 < b_6 \le b_4 \le b_3 \le b_2/2$, we have $x_0 \in B(x^*, b_2/2)$. Therefore, from Lemma 5.3, we obtain

$$\begin{aligned}
\|x_1 - x^*\| &= \|x_0 + t_0 d_0 - x^*\| \\
&\le \|x_0 - x^*\| + \|d_0\| \\
&\le \|x_0 - x^*\| + \kappa_2 \mathrm{dist}(x_0, X^*) \\
&\le (1 + \kappa_2)\|x_0 - x^*\| \\
&\le (1 + \kappa_2)b_7 \\
&\le \frac{1 + \kappa_2}{1 + 2\kappa_2}b_6 \le b_6,
\end{aligned}$$

which shows that $x_1 \in B(x^*, b_6)$. Next, we consider the case where $k \ge 1$. Suppose that $x_j \in B(x^*, b_6),\ j = 1, \ldots, k$. It follows from Lemma 5.6 that

$$\mathrm{dist}(x_j, X^*) \le \frac{1}{2}\mathrm{dist}(x_{j-1}, X^*) \le \cdots \le \left(\frac{1}{2}\right)^j \mathrm{dist}(x_0, X^*) \le \left(\frac{1}{2}\right)^j \|x_0 - x^*\| \le \left(\frac{1}{2}\right)^j b_7.$$

Therefore,

$$\|d_j\| \le \kappa_2 \mathrm{dist}(x_j, X^*) \le \left(\frac{1}{2}\right)^j \kappa_2 b_7. \qquad (5.12)$$

Thus we obtain

$$\|x_{k+1} - x^*\| \le \|x_0 - x^*\| + \sum_{j=0}^{k}\|d_j\| \le (1 + 2\kappa_2)b_7 = b_6,$$

which shows that $x_{k+1} \in B(x^*, b_6)$. This completes the proof. □

By using Lemmas 5.6 and 5.7, we give the rate of convergence of the E-RNM.

Theorem 5.1. *Suppose that Assumption 3 holds. Let* $\{x_k\}$ *be a sequence generated by the E-RNM with* $x_0 \in B(x^*, b_7)$. *Then,* $\{\mathrm{dist}(x_k, X^*)\}$ *converges to* 0 *at the rate of* $1+\delta$. *Moreover,* $\{x_k\}$ *converges to a local optimal solution* $\hat{x} \in B(x^*, b_6)$.

Proof. The first part of the theorem directly follows from Lemmas 5.6 and 5.7. So we only show the second part. For all integers $p > q \geq 0$, we obtain

$$\begin{aligned}\|x_p - x_q\| &\leq \sum_{j=q}^{p-1} \|d_j\| \\ &\leq \kappa_2 b_7 \sum_{j=q}^{p-1} \left(\frac{1}{2}\right)^j \\ &\leq \kappa_2 b_7 \sum_{j=q}^{\infty} \left(\frac{1}{2}\right)^j \\ &\leq \kappa_2 b_7 \left(\frac{1}{2}\right)^{q-1},\end{aligned}$$

where the second inequality follows from (5.12). Thus $\{x_k\}$ is a Cauchy sequence, and hence it converges. □

Remark 5.1. *Note that by using techniques similar to [3], we can prove that* $\|x_k - \hat{x}\|$ *converges to* 0 *at the rate of* $1+\delta$.

6 Concluding Remarks

In this paper, we have considered the RNM extended to the unconstrained nonconvex optimization. We have shown that the E-RNM has a global convergence and a superlinear convergence under appropriate conditions. Moreover, we have shown that the global complexity bound of the E-RNM is $O(\epsilon^{-2})$ when $\nabla^2 f$ is Lipschitz continuous. To our knowledge, this complexity is best for RNMs.

For future work, we may consider to improve the global complexity bound of the E-RNM when f is convex. Moreover, it would be important to investigate how to calculate a search direction d_k efficiently or how to choose the parameters δ, c_1, c_2.

References

[1] H. Dan, N. Yamashita, and M. Fukushima, *Convergence properties of the inexact Levenberg-Marquardt method under local error bound conditions*, Optimization Methods and Software, 17 (2002), pp. 605–626.

[2] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Canbridge University Press, 1985.

[3] D. H. Li, M. Fukushima, L. Qi, and N.Yamashita, *Regularized Newton methods for convex minimization problems with singular solutions*, Computational Optimization and Applications, 28 (2004), pp. 131–147.

[4] R. A. Polyak, *Regularized Newton method for unconstrained convex optimization*, to appear in Mathematical Programming. DOI 10.1007/s 10107-007-0143-3, published online 06/16/2007.

[5] N. Yamashita and M. Fukushima, *On the rate of convergence of the Levenberg-Marquardt method*, Computing. Supplementum(Wien), 15 (2001), pp. 227–238.