A Regularized Newton Method with Correction for Unconstrained Nonconvex Optimization

Heng Wang[1] & Mei Qin[1]

[1] College of Science, University of Shanghai for Science and Technology, Shanghai, China

Correspondence: Heng Wang, College of Science, University of Shanghai for Science and Technology, Shanghai, China 200093. E-mail: wanghengusst@126.com

Abstract

In this paper, we present a modified regularized Newton method for minimizing a nonconvex function whose Hessian matrix may be singular. We show that if the gradient and Hessian of the objective function are Lipschitz continuous, then the method has a global convergence property. Under the local error bound condition which is weaker than nonsingularity, the method has cubic convergence.

Keywords: regularized Newton method, trust region method, nonconvex optimization, local error bound, cubic convergence

1. Introduction

We consider the unconstrained optimization problem

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

where $f : R^n \to R$ is twice continuously differentiable, whose gradient ∇f and Hessian $\nabla^2 f$ are denoted by $g(x)$ and $H(x)$ respectively. Throughout this paper, we assume that the solution set of (1) 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 system of nonlinear equations

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

Hence, we could get the minimizer of $f(x)$ by solving (2) (C.T.Kelley, 1999, W. Sun, 2006, W. Zhou, 2008). The Newton method is one of a efficient solution method. At every iteration, it computes the trial step

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

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.

To overcome the difficulty caused by the possible singularity of H_k, (D. Sun, 1999) proposed a regularized Newton method, where the trial step is the solution of the linear equations

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

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

Now we need to consider another question, "how to choose the modified regularized parameter μ_k? " which will play important roles not only in theoretical analysis but also in numerical experiments. Yamashita and Fukushima (D. H. Li, 2004) chose $\mu_k = \|g_k\|^2$ and showed that the regularized Newton method has quadratic convergence under the local error bound condition which is weaker than nonsingularity. Fan and Yuan (J. Y. Fan, 2005) took $\mu_k = \|g_k\|^\delta$ with $\delta \in [1, 2]$ and showed that the Levenberg-Marqularity method preserves the quadratic convergence under the same conditions. Numerical results (J. Y. Fan, 2009, Jinyan Fan, 2014) show that the choice of $\mu_k = \|F_k\|$ performs more stable and preferable.

In most past studies (N. Yamashita, 2001, Polyak, 2009, J. Y. Fan, 2009, Jinyan Fan, 2014) for the regularized Newton method, the convergence properties have been discussed only when f is convex. In this paper, we propose a modified Newton method for (1) whose objective function f is nonconvex, which is mainly motivated in (Dong-huiLi, 2004). Dong-Huili, Masao Fukushima, Liqun Qi and Nobuo Yamashita have proposed that regularized Newton and inexact Newton methods are possible extended to nonconvex minimization problems (Dong-huiLi, 2004). They chose Λ_k to satisfy

$$\Lambda_k = \max(C\,\|g_k\|, -2\lambda_1(H_k)), \tag{5}$$

where $C > 0$ is a constant and $\lambda_1(H_k)$ is the minimum eigenvalue of H_k. Based on the better performance of the modified regularized method with $\Lambda_k = \max(0, -\lambda_{\min}(H_k))$, we will consider the choice of

$$\Lambda_k = \max(0, -\lambda_{\min}(H_k)), \tag{6}$$

in this paper.

We extend the regularized Newton method (4) to the unconstrained nonconvex optimization. At the k-th iteration of the modified regularized Newton method , we set regularized parameter μ_k as $\mu_k = \alpha_1 \Lambda_k + \|g_k\|$, where $\alpha_1 \geq 1$. From the definition of Λ_k, the matrix $H_k + \alpha_1 \Lambda_k I$ is positive semidefinite even if f is nonconvex. Therefore, if $\|g_k\| \neq 0$, then $H_k + \mu_k I = H_k + \alpha_1 \Lambda_k I + \|g_k\| I > 0$, we can use regularized Newton method to solve the problem of (1).

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

$$(H_k + \mu_k I)d = -g_k \tag{7}$$

to obtain the Newton step d_k, where $\mu_k = \alpha_1 \Lambda_k + \lambda_k \|g_k\|$, and then solves the linear equations

$$(H_k + \mu_k I)d = -g(y_k) \ \ with \ y_k = x_k + d_k \tag{8}$$

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

The paper is organized as follows. In section 2, we present a new modified regularized Newton algorithm by using trust region technique, then prove the global convergence. In section 3, we study the convergence rate of the algorithm and obtain the cubic convergence under the local error bound condition. Finally, we conclude the paper in section 4.

2. The Algorithm and Global Convergence

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

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

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

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

If we let

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

then d_k is also a solution of the trust region problem

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

By the famous result given by Powell in (M.J.D.Powell, 1975), we know that

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

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

$$\min_{d \in R^n} \frac{1}{2} d^T H_k d + g(y_k)^T d + \frac{1}{2} \mu_k \|d\|^2,$$

but also the solution of the following trust region problem

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

where

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

Therefore we also have

$$\phi(0) - \phi(\widetilde{d_k}) \geq \frac{1}{2} \|g(y_k)\| \min \left\{ \left\| \widetilde{d_k} \right\|, \frac{\|g(y_k)\|}{\|H_k\|} \right\}. \tag{11}$$

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

$$\text{Pr} ed_k = \varphi(0) - \varphi(d_k) + \phi(0) - \phi\left(\widetilde{d_k}\right), \tag{12}$$

which satisfies

$$\text{Pr} ed_k \geq \frac{1}{2} \|g_k\| \min \left\{ \|d_k\|, \frac{\|g_k\|}{\|H_k\|} \right\} + \frac{1}{2} \|g(y_k)\| \min \left\{ \left\| \widetilde{d_k} \right\|, \frac{\|g(y_k)\|}{\|H_k\|} \right\}. \tag{13}$$

The ratio of the actual reduction to the predicted reduction

$$r_k = \frac{Ared_k}{\text{Pr} ed_k}, \tag{14}$$

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

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

Algorithm 2.1

Step 1. Given $x_0 \in R^n$, $\varepsilon \geq 0$, $\lambda_0 > m > 0$, $0 < c_0 \leq c_1 \leq c_2 < 1$, $0 < p_1 < 1 < p_2$, $\alpha_1 \geq 1$, $k := 0$.
Step 2. If $\|g_k\| \leq \varepsilon$, then stop. Otherwise go to step 3.
Step 3. Compute $\Lambda_k = \max(0, -\lambda_{\min}(H_k))$, $\quad \nu_k = \lambda_k \|g_k\|$.

    Solve

$$(H_k + \alpha_1 \Lambda_k I + \nu_k I) d = -g_k. \tag{15}$$

    to obtain d_k.

    Set

$$y_k = x_k + d_k.$$

    Solve

$$(H_k + \alpha_1 \Lambda_k I + \nu_k I) d = -g(y_k), \tag{16}$$

    to obtain $\widetilde{d_k}$.

    Set

$$s_k = d_k + \widetilde{d_k}.$$

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

    Set

$$x_{k+1} = \begin{cases} x_k + s_k, & if \ r_k \geq c_0, \\ x_k, & otherwise. \end{cases} \tag{17}$$

Step 5. Update λ_{k+1} as

$$\lambda_{k+1} = \begin{cases} p_2 \lambda_k, & if \ r_k < c_1, \\ \lambda_k, & if \ r_k \in [c_1, c_2], \\ \max\{p_1 \lambda_k, m\}, & if \ r_k > c_2. \end{cases} \tag{18}$$

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

$$\|g(y) - g(x)\| \le L_1 \|y - x\|, \ \forall x, y \in R^n \tag{19}$$

and

$$\|H(y) - H(x)\| \le L_2 \|y - x\|, \ \forall x, y \in R^n. \tag{20}$$

It follows from (20) that

$$\|g(y) - g(x) - H(x)(y - x)\| \le L_2 \|y - x\|^2, \ \forall x, y \in R^n. \tag{21}$$

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

Lemma 2.1 A real-valued matrix is positive semidefinite if and only if $\left(A + A^T\right)/2$ is positive semidefinite.

Proof. See (W. Sun, 2006). □

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

Lemma 2.2 Suppose A is symmetric positive semidefinite. Then,

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

and

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

hold for any $\varphi > 0$.

Proof. See (JinyanFan, 2014). □

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

$$\liminf_{k \to \infty} \|g_k\| = 0. \tag{22}$$

Proof. The proof is similar to (Weijun Zhou, 2013). We prove by contradiction. If the theorem is not true, then there exists a positive τ and an integer $\widetilde{k}$ such that

$$\|g_k\| \ge \tau, \ \forall k \ge \widetilde{k}. \tag{23}$$

Without loss of generality, we can suppose $\widetilde{k} = 1$. Set $T = \{k | x_k \ne x_{k+1}\}$. 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 (17), we have

$$r_k < c_0, \ \forall k \ge k_1.$$

Therefore by (18) and (23), we deduce

$$\lambda_k \to \infty, \ \nu_k \to \infty. \tag{24}$$

Since $x_{k+1} = x_k, \forall k \geq k_1$, we get from (15) and (24) that

$$\|d_k\| = \left\| -(H_k + \alpha_1 \Lambda_k I + \lambda_k \|g_k\| I)^{-1} g_k \right\| \to 0. \tag{25}$$

From (16), we obtain

$$
\begin{aligned}
\left\| \widetilde{d_k} \right\| &= \left\| -(H_k + \alpha_1 \Lambda_k I + \lambda_k \|g_k\| I)^{-1} g(y_k) \right\| \\
&\leq \left\| (H_k + \alpha_1 \Lambda_k I + \lambda_k \|g_k\| I)^{-1} (g(y_k) - g_k - H_k d_k) \right\| \\
&\quad + \left\| (H_k + \alpha_1 \Lambda_k I + \lambda_k \|g_k\| I)^{-1} g_k \right\| + \left\| (H_k + \alpha_1 \Lambda_k I + \lambda_k \|g_k\| I)^{-1} H_k d_k \right\| \\
&\leq L_2 \lambda_k^{-1} \frac{1}{\|g_k\|} \|d_k\|^2 + 2 \|d_k\| \\
&\leq \gamma_1 \|d_k\|,
\end{aligned}
\tag{26}
$$

where γ_1 is a positive constant.

It follows from (9) and (12) that

$$
\begin{aligned}
|Ared_k - \mathrm{Pr}\,ed_k| &= \left| f(x_k) - f(x_k + d_k + \widetilde{d_k}) - (\phi_{k,1}(0) - \phi_{k,1}(d_k) + \phi_{k,2}(0) - \phi_{k,2}(\widetilde{d_k})) \right| \\
&\leq \left| f(y_k + \widetilde{d_k}) - f(y_k) - \frac{1}{2} \widetilde{d_k}^T H_k \widetilde{d_k} - g(y_k)^T \widetilde{d_k} \right| \\
&\quad + \left| f(y_k) - f(x_k) - \frac{1}{2} d_k^T H_k d_k - g_k^T d_k \right| \\
&\leq o\left(\|d_k\|^2 \right) + o\left(\left\| \widetilde{d_k} \right\|^2 \right).
\end{aligned}
\tag{27}
$$

Moreover, from (13), (23), (19) and (25), we have

$$\mathrm{Pr}\,ed_k \geq \frac{1}{2} \tau \min \left\{ \|d_k\|, \frac{\tau}{L_1} \right\} \geq \frac{1}{2} \tau \|d_k\|, \tag{28}$$

for sufficiently large k.

Duo to (27) and (28), we get

$$
\begin{aligned}
|r_k - 1| &= \left| \frac{Ared_k - \mathrm{Pr}\,ed_k}{\mathrm{Pr}\,ed_k} \right| \\
&\leq \left| \frac{f(x_k) - f(x_k + d_k + \widetilde{d_k}) - (\phi_{k,1}(0) - \phi_{k,1}(d_k) + \phi_{k,2}(0) - \phi_{k,2}(\widetilde{d_k}))}{\frac{1}{2} \tau \min \left\{ \|d_k\|, \frac{\tau}{L_1} \right\}} \right| \\
&\leq \frac{o\left(\|d_k\|^2 \right) + o\left(\left\| \widetilde{d_k} \right\|^2 \right)}{\|d_k\|} \to 0,
\end{aligned}
\tag{29}
$$

which implies that $r_k \to 1$. Hence, there exists positive constant γ_2 such that $\lambda_k \leq \gamma_2$, holds for all large k, which contradicts to (24).

Case (2): T is infinite. Then we have from (13) and (23) that

$$
\begin{aligned}
\infty > f(x_1) - \liminf_{k \to \infty} f(x_k) &\geq \sum_{i=1}^{\infty} (f(x_i) - f(x_{i+1})) \\
&= \sum_{k \in T} (f(x_k) - f(x_{k+1})) \geq \sum_{k \in T} c_0 \mathrm{Pr}\,ed_k \\
&\geq \sum_{k \in T} c_0 \left(\frac{1}{2} \|g_k\| \min \left\{ \|d_k\|, \frac{\|g_k\|}{\|H_k\|} \right\} + \frac{1}{2} \|g(y_k)\| \min \left\{ \left\| \widetilde{d_k} \right\|, \frac{\|g(y_k)\|}{\|H_k\|} \right\} \right) \\
&\geq \sum_{k \in T} c_0 \frac{\tau}{2} \min \left\{ \|d_k\|, \frac{\tau}{L_1} \right\},
\end{aligned}
\tag{30}
$$

which implies that

$$\lim_{k\to\infty, k\in T} d_k = 0. \tag{31}$$

The above equality together with the updating rule of (18) means

$$\nu_k \to \infty. \tag{32}$$

Similar to (26), it follows from (31) and (32) that

$$\left\| \widetilde{d_k} \right\| \le \gamma_3 \|d_k\|, \quad \forall k \in T$$

for some positive constant γ_3. Then we have

$$\|s_k\| = \left\| d_k + \widetilde{d_k} \right\| \le (1 + \gamma_3) \|d_k\|, \quad \forall k \in T.$$

This equality together with (30) yields

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

which implies that

$$x_k \to x^*. \tag{33}$$

It follows from (15), (33), (32) and (26) that

$$d_k \to 0, \quad \widetilde{d_k} \to 0. \tag{34}$$

Since $(H_k + \alpha_1 \Lambda_k I + \lambda_k \|g_k\| I) d = -g_k$ from (15), we have from (23), (19) and (34) that

$$1 \le \frac{\|H_k\|}{\|g_k\|} \|d_k\| + \frac{\alpha_1 \|\Lambda_k\|}{\|g_k\|} \|d_k\| + \lambda_k \|g_k\| \le \frac{L_1}{\tau} \|d_k\| + \frac{\alpha_1 \|\Lambda_k\|}{\tau} \|d_k\| + \lambda_k \|d_k\|,$$

which means

$$\lambda_k \to \infty. \tag{35}$$

By the same analysis as (29) we know that

$$r_k \to 1. \tag{36}$$

Hence, there exists a positive constant $\gamma_4 > m$ such that $\lambda_k \le \gamma_4$ holds for all sufficiently large k, which gives a contradiction to (35). The proof is completed. $\qquad\square$

3. 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) cubically. To study the local convergence properties of Algorithm 2.1, we make the following assumptions.

Assumption 3.1

(a) The sequence $\{x_k\}$ generated by Algorithm 2.1 converges to $x^* \in X$ and lies in some neighbourhood of x^*.

(b) $\|g(x)\|$ provides a local error bound on some $N(x^*, b_1)$ for (2), that is, there exist constant $\beta_1 > 0$ and $0 < b_1 < 1$ such that

$$\|g(x)\| \ge \beta_1 dist(x, X) \quad \forall x \in N(x^*, b_1) = \{x| \|x - x^*\| \le b_1\}. \tag{37}$$

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

$$\|H(y) - H(x)\| \le \widetilde{L_1} \|y - x\| \quad \forall x, y \in N(x^*, b_1). \tag{38}$$

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 (N.Yamashita, 2001, Dong-huiLi, 2004). Hence, the local error bound condition is weaker than nonsingularity.

By Assumption 3.1 (c), we know

$$\|g(y) - g(x) - H(x)(y - x)\| \le \widetilde{L_1} \|y - x\|^2 \quad \forall x, y \in N(x^*, b_1) \tag{39}$$

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

$$\|g(y) - g(x)\| \le \widetilde{L_2} \|y - x\| \qquad \forall x, y \in N(x^*, b_1). \tag{40}$$

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 s_k and the distance from x_k to the solution set.

Lemma 3.1 Under the condition of Assumption 3.1, for all sufficiently large k, we have

$$\|s_k\| \le O(dist(x_k, X)) \tag{41}$$

Proof. Since $x_k \to x^* \in X$, we get

$$\|\overline{x_k} - x_k\| = dist(x_k, X) \le \|x_k - x^*\| \to 0.$$

Due to (37), we have

$$v_k = \lambda_k \|g_k\| \ge m\beta_1 dist(x_k, X) = m\beta_1 \|\overline{x_k} - x_k\|.$$

From (15), we get

$$\begin{aligned}
\|d_k\| &= \left\|(H_k + \alpha_1\Lambda_k I + v_k I)^{-1} g_k\right\| \\
&\le \left\|(H_k + \alpha_1\Lambda_k I + v_k I)^{-1} (g(\overline{x_k}) - g_k - H_k(\overline{x_k} - x_k))\right\| \\
&\quad + \left\|(H_k + \alpha_1\Lambda_k I + v_k I)^{-1} H_k(\overline{x_k} - x_k)\right\| \\
&\le \widetilde{L_1} v_k^{-1} \|\overline{x_k} - x_k\|^2 + \|\overline{x_k} - x_k\| \\
&\le O(\|\overline{x_k} - x_k\|).
\end{aligned} \tag{42}$$

Since $y_k = x_k + d_k$, then $y_k \to x^*$, which means $y_k \in N(x^*, b_1)$ for sufficiently large k.
From (16), we get

$$\begin{aligned}
\left\|\widetilde{d_k}\right\| &= \left\|(H_k + \alpha_1\Lambda_k I + v_k I)^{-1} g(y_k)\right\| \\
&\le \left\|(H_k + \alpha_1\Lambda_k I + v_k I)^{-1} (g(y_k) - g_k - H_k d_k)\right\| \\
&\quad + \left\|(H_k + \alpha_1\Lambda_k I + v_k I)^{-1} g_k\right\| + \left\|(H_k + \alpha_1\Lambda_k I + v_k I)^{-1} H_k d_k\right\| \\
&\le \widetilde{L_1} v_k^{-1} \|d_k\|^2 + 2\|d_k\| \\
&\le O(\|\overline{x_k} - x_k\|).
\end{aligned} \tag{43}$$

Combining (42) and (43), we obtain

$$\|s_k\| = \left\|d_k + \widetilde{d_k}\right\| \le \|d_k\| + \left\|\widetilde{d_k}\right\| \le O(\|\overline{x_k} - x_k\|). \tag{44}$$

$\square$

3.1 The Boundedness of λ_k and Λ_k

In the following, we will show λ_k and Λ_k are bounded above, which will play a key role in the next subsection.

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

$$\lambda_k \le T$$

holds for all sufficiently large k.

Proof. From (10), (37) and (40), we have

$$\begin{aligned}
\varphi(0) - \varphi(d_k) &\ge \frac{1}{2} \|g_k\| \min\left\{\|d_k\|, \frac{\|g_k\|}{\|H_k\|}\right\} \\
&\ge \frac{1}{2}\beta_1 \|\overline{x_k} - x_k\| \min\left\{\|d_k\|, \frac{\beta_1 \|\overline{x_k} - x_k\|}{\widetilde{L_2}}\right\} \\
&\ge \beta_2 \|\overline{x_k} - x_k\| \min\{\|d_k\|, \|\overline{x_k} - x_k\|\}
\end{aligned} \tag{45}$$

for some constant β_2.

Then from (13), (27), (42), (43) and (45), we get

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

which implies that $r_k \to 1$. Therefore there exists a constant $T > m$ such that $\lambda_k \le T$ holds for all sufficiently large k. The proof is completed. $\qquad\square$

Lemma 3.3 Suppose Assumption 3.1 hold. If $x_k \in N(x^*, b_1/2)$, then

$$
\Lambda_k \le \widetilde{L_1} dist(x_k, X).
$$

Proof. Using a method similar to the proof of (Kenji Ueda, 2010), we get the expected result. $\qquad\square$

3.2 Cubic Convergence of Algorithm 2.1

Lemma 3.4 Suppose Assumption 3.1 hold, then we have

$$
dist(x_{k+1}, X) \le O(dist(x_k, X)^2). \tag{46}
$$

Proof. From (37) and (40), we have

$$
m\beta_1 \|\overline{x_k} - x_k\| \le v_k = \lambda_k \|g_k\| = \lambda_k \|g_k - g(\overline{x_k})\| \le T\widetilde{L_2} \|\overline{x_k} - x_k\|. \tag{47}
$$

Which shows that $\|\overline{x_k} - x_k\|$ is equivalent to v_k.

From the local error bound condition, (39), (16) and (38), we have

$$
\begin{aligned}
\beta_1 \|\overline{x_{k+1}} - x_{k+1}\| \le \|g(x_{k+1})\| &= \left\| g\left(y_k + \widetilde{d_k}\right) \right\| \\
&\le \left\| g\left(y_k + \widetilde{d_k}\right) - g(y_k) - H(y_k)\widetilde{d_k} \right\| + \left\| g(y_k) + H(y_k)\widetilde{d_k} \right\| \\
&\le \widetilde{L_1} \left\|\widetilde{d_k}\right\|^2 + \widetilde{L_1} \|d_k\| \left\|\widetilde{d_k}\right\| + (\alpha_1 \Lambda_k + v_k) \left\|\widetilde{d_k}\right\| \\
&\le O\left(\|\overline{x_k} - x_k\|^2\right).
\end{aligned} \tag{48}
$$

$\qquad\square$

Note that since

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

we may deduce from (48) that

$$
\|\overline{x_k} - x_k\| \le 2 \|s_k\| \tag{49}
$$

for all sufficiently large k. Combining this inequality with (44) and (48), we obtain that

$$
\|s_{k+1}\| = O\left(\|s_k\|^2\right), \tag{50}
$$

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

$$
\|x_{k+1} - x^*\| = O\left(\|x_k - x^*\|^2\right). \tag{51}
$$

To obtain faster convergence of the modified regularized Newton method, we need to estimate $\left\|\widetilde{d_k}\right\|$ more accurately. We will use the SVD technique to derive the local convergence rate of Algorithm 2.1. Since $H(x^*)$ is symmetric, there is an orthogonal matrix $(V_{k,1}^*, V_{k,2}^*)$ such that

$$H(x^*) = \left(V_{k,1}^*, V_{k,2}^*\right) \begin{pmatrix} \Sigma_{k,1}^* & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{k,1}^{*T} \\ V_{k,2}^{*T} \end{pmatrix} = V_{k,1}^* \Sigma_{k,1}^* V_{k,1}^{*T},$$

where $\Sigma_{k,1}^*$ is a diagonal matrix. Moreover, we can suppose that $H(x)$ has the following decomposition

$$H(x) = \left(V_{k,1}, \ V_{k,2}\right) \begin{pmatrix} \Sigma_{k,1} & 0 \\ 0 & \Sigma_{k,2} \end{pmatrix} \begin{pmatrix} V_{k,1}{}^T \\ V_{k,2}{}^T \end{pmatrix} = V_{k,1}\Sigma_{k,1}V_{k,1}{}^T + V_{k,2}\Sigma_{k,2}V_{k,2}{}^T,$$

where $rank\left(\Sigma_{k,1}\right) = rank\left(\Sigma_{k,1}^*\right)$ and $\Sigma_{k,2}$ converges to zero as $x \to x^*$. In the following, we neglect the subscription k in $\Sigma_{k,i}$ and $V_{k,i}$ $(i = 1, 2)$, and write $H(x_k)$ as

$$H(x_k) = V_1\Sigma_1 V_1{}^T + V_2\Sigma_2 V_2{}^T.$$

Lemma 3.5 (W. Sun, 2006). If the sequence $\{x_k\}$ converges superlinearly to x^*, then

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} = 1.$$

Proof. See (W. Sun, 2006). $\qquad\square$

Therefore we have from Lemma3.5, (50), (51) and (44) that there exists two positive constants β_3 and β_4 such that

$$\|x_k - x^*\| \le \beta_3 \|s_k\| \le \beta_4 \|\overline{x_k} - x_k\| \le \beta_4 \|x_k - x^*\|, \tag{52}$$

which means that $\|x_k - x^*\|$ is equivalent to $\|\overline{x_k} - x_k\|$.

By the theory of matrix perturbation (G.W.Stewart, 1990) and (38), we have

$$\left\|\Sigma_1 - \Sigma_1^*\right\| + \|\Sigma_2\| \le \|H_k - H(x^*)\| \le \widetilde{L_1} \|x_k - x^*\|.$$

Combining this inequality with (52), we get

$$\left\|\Sigma_1 - \Sigma_1^*\right\| \le \widetilde{L_1} \|\overline{x_k} - x_k\|, \quad \|\Sigma_2\| \le \widetilde{L_1} \|\overline{x_k} - x_k\|.$$

Lemma 3.6 Suppose Assumption 3.1 hold, then we have

$$\|g(y_k)\| \le O\left(\|\overline{x_k} - x_k\|^2\right),$$

$$\left\|U_2 U_2^T g(y_k)\right\| \le O\left(\|\overline{x_k} - x_k\|^3\right).$$

Proof. From lemma 3.3, (15), (42) and (47), we have

$$\|g_k + H_k d_k\| = (\alpha_1\Lambda_k + v_k) \|d_k\| \le O\left(\|\overline{x_k} - x_k\|^2\right). \tag{53}$$

Similarly, we know

$$\left\|g(y_k) + H_k\widetilde{d_k}\right\| = (\alpha_1\Lambda_k + v_k) \left\|\widetilde{d_k}\right\| \le O\left(\|\overline{x_k} - x_k\|^2\right). \tag{54}$$

Then we get from (39), (42) and (53) that

$$\|g(y_k)\| = \|g(x_k + d_k) - g_k - H_k d_k\| + \|g_k + H_k d_k\| \\ \le O\left(\|\overline{x_k} - x_k\|^2\right). \tag{55}$$

From the local error bound condition and (55), we have

$$\|\overline{y_k} - y_k\| \le \beta_1^{-1} \|g(y_k)\| \le O\left(\|\overline{x_k} - x_k\|^2\right).$$

Set $\widetilde{H_k} = U_1\Sigma_1 U_1^T$ and $\widetilde{d_k} = -\widetilde{H_k}^+ g(y_k)$, then $\widetilde{d_k}$ is the least square solution of $\min \left\| g(y_k) + \widetilde{H_k}d \right\|$. Therefore we have

$$
\begin{aligned}
\left\| U_2 U_2^T g(y_k) \right\| = \left\| g(y_k) + \widetilde{H_k}\widetilde{d_k} \right\| &\le \left\| g(y_k) + \widetilde{H_k}(\overline{y_k} - y_k) \right\| \\
&\le \| g(y_k) + H(y_k)(\overline{y_k} - y_k)\| + \left\| \left(\widetilde{H_k} - H(y_k)\right)(\overline{y_k} - y_k) \right\| \\
&\le \| g(y_k) + H(y_k)(\overline{y_k} - y_k)\| + \|(H(y_k) - H_k)(\overline{y_k} - y_k)\| \\
&\quad + \left\| \left(H_k - \widetilde{H_k}\right)(\overline{y_k} - y_k) \right\| \\
&\le \widetilde{L_1}\|\overline{y_k} - y_k\|^2 + \widetilde{L_1}\|d_k\|\|\overline{y_k} - y_k\| + \|\Sigma_2\|\|\overline{y_k} - y_k\| \\
&\le O\left(\|\overline{x_k} - x_k\|^3\right).
\end{aligned}
\tag{56}
$$

$\square$

Lemma 3.7 Suppose Assumption 3.1 hold, then we have

$$
\left\| \widetilde{d_k} \right\| \le O\left(\|\overline{x_k} - x_k\|^2\right).
\tag{57}
$$

Proof. From (16), we have

$$
\begin{aligned}
\widetilde{d_k} &= -(H_k + \alpha_1\Lambda_k I + v_k I)^{-1} g(y_k) \\
&= -U_1(\Sigma_1 + \mu_k I)^{-1} U_1^T g(y_k) - U_2(\Sigma_2 + \mu_k I)^{-1} U_2^T g(y_k).
\end{aligned}
\tag{58}
$$

Since $x_k \to x^*$, then $\Sigma_1 \to \Sigma_1^*$ and hence Σ_1^{-1} is uniformly bounded, that is, there exists a constant β_5 such that

$$
\left\| \Sigma_1^{-1} \right\| \le \beta_5.
\tag{59}
$$

Then from (58), (59) and lemma 3.6, we obtain

$$
\begin{aligned}
\left\| \widetilde{d_k} \right\| &\le \left\| \Sigma_1^{-1} \right\| \left\| U_1 U_1^T g(y_k) \right\| + \mu_k^{-1} \left\| U_2 U_2^T g(y_k) \right\| \\
&\le \beta_5 \| g(y_k)\| + \mu_k^{-1} \left\| U_2 U_2^T g(y_k) \right\| \\
&\le O\left(\|\overline{x_k} - x_k\|^2\right).
\end{aligned}
\tag{60}
$$

$\square$

Theorem 3.1 Suppose Assumption 3.1 hold, the sequence generated by Algorithm 2.1 converges to some solution of (1) cubically.

Proof. It follows from (37), we get

$$
\begin{aligned}
c_1 \|\overline{x_{k+1}} - x_{k+1}\| \le \| g(x_{k+1})\| = \left\| g\left(y_k + \widetilde{d_k}\right) \right\| &\le \left\| g\left(y_k + \widetilde{d_k}\right) - g(y_k) - H(y_k)\widetilde{d_k} \right\| + \left\| g(y_k) + H(y_k)\widetilde{d_k} \right\| \\
&\le \widetilde{L_1}\left\| \widetilde{d_k} \right\|^2 + \left\| g(y_k) + H_k\widetilde{d_k} \right\| + \left\| (H(y_k) - H_k)\widetilde{d_k} \right\| \\
&\le \widetilde{L_1}\left\| \widetilde{d_k} \right\|^2 + \widetilde{L_1}\|d_k\|\left\| \widetilde{d_k} \right\| + (\alpha_1\Lambda_k + v_k)\left\| \widetilde{d_k} \right\| \\
&\le O\left(\|\overline{x_k} - x_k\|^3\right),
\end{aligned}
\tag{61}
$$

From Lemma 3.1, (49) and (61), we finally have

$$
\|s_{k+1}\| = O\left(\|s_k\|^3\right),
\tag{62}
$$

which indicates that $\{x_k\}$ converges to x^* cubically, i.e., $\|x_{k+1} - x^*\| = O\left(\|x_k - x^*\|^3\right)$. The proof is completed. $\square$

4. Concluding Remarks

In this paper, we propose a modified regularized Newton method with correction for unconstrained nonconvex optimization. Furthermore, we have proved that the modified regularized Newton method has a global convergence and a local cubic convergence under some appropriate conditions.

Acknowledgements

References

Dong, H. L., Masao, F., Liqun, Q., & Nobuo, Y. (2004). Regularized Newton methods for convex minimization problems with singular solutions. *Computational Optimization and Applications, 28*, 131-147. http://dx.doi.org/10.1023/B:COAP.0000026881.96694.32

Fan, J. Y., & Pan, J. Y. (2009). A note on the Levenberge-Marquardt parameter. *Appl. Math. Comput., 207*, 351-359. http://dx.doi.org/10.1016/j.amc.2008.10.056

Fan, J. Y., & Yuan, Y. X. (2005). On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. *Computing, 74*, 23-39. http://dx.doi.org/10.1007/s00607-004-0083-1

Fan, J. Y., & Yuan, Y. X. (2014). A regularized Newton method for monotone nonlinear equations and its application. *Optimization Methods and Software, 29*, 102-119. http://dx.doi.org/10.1080/10556788.2012.746344

Horn, R. A., & Johnson, C. R. (1985). Matrix analysis,Cambridge University press, Cambridge.

Kelley, C. T. (1999). Iterative Methods for Optimization. *Frontiers in Applied Mathematics, 18*, SIAM, Philadelphia.

Kenji, U. & Nobuo, Y. (2010). Convergence Properties of the Regularized Newton Method for the Unconstrained Nonconvex Optimization. *Appl Math Optim., 62*, 27-46. http://dx.doi.org/10.1007/s00245-009-9094-9

Polyak, R. A. (2009). Regularized Newton method for unconstrained convex optimization. *Math. Program., Ser B 120*, 125-145. http://dx.doi.org/10.1007/s10107-007-0143-3

Powell, M. J. D. (1975). 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.

Stewart, G. W., & Sun, J. G. (1990). Matrix Perturbation Theory,Academic Press, San Diego, CA.

Sun, D. (1999). A regularization Newton method for solving nonlinear complementarity problems. *Appl. Math. Optim., 40*, 315-339. http://dx.doi.org/10.1007/s002459900128

Sun, W., & Yuan, Y. (2006). *Optimization Theory and Methods*. New York, NY: Springer Science and Business Media. http://dx.doi.org/10.1007/b106451

Weijun, Z., & Xinlong, C. (2013). On the convergence of a modified regularized Newton method for convex optimization with singular solutions. *Journal of Computational and Applied Mathematics, 239*, 179-188. http://dx.doi.org/doi:10.1016/j.cam.2012.09.030

Yamashita, N., & Fukushima, M. (2001). *On the rate of convergence of the Levenberg-Marguardt method, Computing, 15*, 227-238.

Zhou, W., & Li, D. (2008). A globally convergent BFGS method for nonlinear monotone equations without any merit functions. *Math. Comp., 77*, 2231-2240.

Copyrights