Journal of Computational and Applied Mathematics 239 (2013) 179–188

Contents lists available at SciVerse ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

On the convergence of a modified regularized Newton method for convex optimization with singular solutions

Weijun Zhou *, Xinlong Chen

Department of Mathematics, Changsha University of Science and Technology, Changsha 410004, China

ARTICLE INFO

Article history:
Received 21 November 2011
Received in revised form 21 September 2012

MSC:
65K05
90C30

Keywords:
Convex optimization
Newton method
Global convergence
Cubic convergence

ABSTRACT

In this paper we propose a modified regularized Newton method for convex minimization problems whose Hessian matrices may be singular. The proposed method is proved to converge globally if the gradient and Hessian of the objective function are Lipschitz continuous. Under the local error bound condition, we first show that the method converges quadratically, which implies that $\|x_k - x^*\|$ is equivalent to $\mathrm{dist}(x_k, X)$, where X is the solution set and $x_k \to x^* \in X$. Then we in turn prove the cubic convergence of the proposed method under the same local error bound condition, which is weaker than nonsingularity.

1. Introduction

Consider the unconstrained optimization problem

$$\min f(x), \quad x \in R^n, \tag{1.1}$$

where $f : R^n \to R$ is convex and twice continuously differentiable, whose gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$ are denoted by $g(x)$ and $G(x)$, respectively. Throughout the paper, we suppose that the solution set X of (1.1) is nonempty, and in all cases $\|\cdot\|$ stands for the 2-norm. It is clear that X is a closed convex set.

It is well-known that $f(x)$ is convex if and only if $G(x)$ is positive semidefinite for all $x \in R^n$. Moreover, if f is convex, then $x \in X$ if and only if x is a solution of the nonlinear equations

$$g(x) = 0. \tag{1.2}$$

There are many efficient methods [1–4] for solving the problem (1.1) or (1.2). The Newton method is one of the best known methods. At each iteration, the Newton method computes the trial step

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

where $g_k = g(x_k)$ and $G_k = G(x_k)$. An attractive feature of the Newton method is that it possesses quadratic convergence rate if $G(x^*)$ is nonsingular at a solution x^*, which implies that the solution is locally isolated.

However, the condition on the nonsingularity of the Hessian is too strong since many problems have singular solutions [5–7], which may contain some inverse problems and ill-posed problems [8]. To obtain reasonable solutions for

* Corresponding author. Tel.: +86 07318828171; fax: +86 07318823826.
E-mail address: weijunzhou@126.com (W. Zhou).

doi:10.1016/j.cam.2012.09.030

this kind of problems, different regularization techniques are often used. Recently, under the local error bound condition, which is weaker than nonsingularity, Li et al. [6] proposed a regularized Newton method with quadratic convergence, where the trial step is the solution of the linear equations

$$(G_k + \lambda_k I)d = -g_k \quad \text{with } \lambda_k = C\|g_k\|$$

for some positive constant C, where I is the identity matrix. More details on the local error bound condition for nonlinear equations can be found in [5,7,9–12].

In this paper we propose a modified regularized Newton method for (1.1), which is mainly motivated in [9], where a modified Levenberg–Marquardt method was proposed for nonlinear equations with cubic convergence under the local error bound condition.

The main scheme of the modified regularized Newton method is given as follows. At each iteration, it solves the linear equations

$$(G_k + \lambda_k I)d = -g_k \tag{1.3}$$

to obtain the Newton step d_k, where λ_k is a suitable regularized parameter, and then solves the linear equations

$$(G_k + \lambda_k I)d = -g(y_k) \quad \text{with } y_k = x_k + d_k \tag{1.4}$$

to obtain the approximate Newton step $\hat{d}_k$.

The purpose of this paper is to investigate whether the proposed method has cubic convergence as the modified Levenberg–Marquardt method [9] under the local error bound condition.

The paper is organized as follows. In Section 2, we present the complete modified regularized Newton method carefully. In Section 3, we prove the global convergence of the proposed method. Quadratic convergence and cubic convergence of the proposed method are obtained in Section 4.

2. The algorithm

Let d_k and $\hat{d}_k$ be given by (1.3) and (1.4), respectively. Since the matrix $G_k + \lambda_k I$ is symmetric and positive definite, d_k is a descent direction of $f(x)$ at x_k, but $d_k + \hat{d}_k$ may not be. Hence we use a trust region technique to globalize the proposed method.

Let

$$\mathrm{Ared}_k = f(x_k) - f(x_k + d_k + \hat{d}_k), \tag{2.1}$$

which is called the actual reduction of $f(x)$ at the k-th iteration.

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

$$\min_{d\in R^n} \varphi_{k,1}(d) = \frac{1}{2}d^T G_k d + g_k^T d + \frac{1}{2}\lambda_k\|d\|^2. \tag{2.2}$$

If we let

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

then it can be verified [2, Theorem 6.1.2] that d_k is also a solution of the trust region problem:

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

By the famous result given by Powell in [13] (also see [2, Lemma 6.1.3]), we know that

$$\varphi_{k,1}(0) - \varphi_{k,1}(d_k) \ge \frac{1}{2}\|g_k\| \min\left\{\|d_k\|, \frac{\|g_k\|}{\|G_k\|}\right\}. \tag{2.3}$$

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

$$\min_{d\in R^n} \varphi_{k,2}(d) = \frac{1}{2}d^T G_k d + g(y_k)^T d + \frac{1}{2}\lambda_k\|d\|^2, \tag{2.4}$$

but also the solution of the following trust region problem:

$$\min_{d\in R^n} \frac{1}{2}d^T G_k d + g(y_k)^T d, \quad \text{s.t. } \|d\| \le \Delta_{k,2},$$

where

$$\Delta_{k,2} = \|\hat{d}_k\| = \| -(G_k + \lambda_k I)^{-1} g(y_k)\|.$$

Therefore we also have

$$\varphi_{k,2}(0) - \varphi_{k,2}(\hat{d}_k) \geq \frac{1}{2}\|g(y_k)\| \min\left\{\|\hat{d}_k\|, \frac{\|g(y_k)\|}{\|G_k\|}\right\}. \tag{2.5}$$

Then we define prediction reduction as

$$\mathrm{Pred}_k = \varphi_{k,1}(0) - \varphi_{k,1}(d_k) + \varphi_{k,2}(0) - \varphi_{k,2}(\hat{d}_k), \tag{2.6}$$

which satisfies

$$\mathrm{Pred}_k \geq \frac{1}{2}\|g_k\| \min\left\{\|d_k\|, \frac{\|g_k\|}{\|G_k\|}\right\} + \frac{1}{2}\|g(y_k)\| \min\left\{\|\hat{d}_k\|, \frac{\|g(y_k)\|}{\|G_k\|}\right\}, \tag{2.7}$$

and it is always nonnegative.

Define the ratio

$$r_k = \frac{\mathrm{Ared}_k}{\mathrm{Pred}_k}, \tag{2.8}$$

which measures the agreement between the model functions and the objective function. Moreover this ratio plays an important role in selecting new iterate x_{k+1} and updating the regularized parameter.

The following is the modified regularized Newton method.

Algorithm 2.1 (*Modified Regularized Newton Algorithm*)**.**

Step 1. Given a starting point $x_1 \in R^n$ and several scalars $\mu_1 > m > 0$, $0 < p_0 \leq p_1 \leq p_2 < 1$. Let $k := 1$.

Step 2. If $\|g_k\| = 0$, then stop. Compute d_k by solving the following linear equations

$$(G_k + \lambda_k I)d = -g_k \quad \text{with } \lambda_k = \mu_k\|g_k\|. \tag{2.9}$$

Set

$$y_k = x_k + d_k. \tag{2.10}$$

Solve

$$(G_k + \lambda_k I)d = -g(y_k) \tag{2.11}$$

to obtain $\hat{d}_k$ and set

$$s_k = d_k + \hat{d}_k. \tag{2.12}$$

Step 3. Compute $r_k = \frac{\mathrm{Ared}_k}{\mathrm{Pred}_k}$. Set

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{if } r_k \geq p_0, \\ x_k, & \text{otherwise.} \end{cases} \tag{2.13}$$

Step 4. Update μ_{k+1} as

$$\mu_{k+1} = \begin{cases} 4\mu_k, & \text{if } r_k < p_1, \\ \mu_k, & \text{if } r_k \in [p_1, p_2], \\ \max\left\{\dfrac{\mu_k}{4}, m\right\}, & \text{if } r_k > p_2. \end{cases} \tag{2.14}$$

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

3. Global convergence

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

Assumption 1. $g(x)$ and $G(x)$ are both Lipschitz continuous, that is, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in R^n \tag{3.1}$$

and

$$\|G(x) - G(y)\| \leq L\|x - y\|, \quad \forall x, y \in R^n. \tag{3.2}$$

It follows from (3.2) that

$$\|g(y) - g(x) - G(x)(y - x)\| \le L\|y - x\|^2, \quad \forall x, y \in R^n. \tag{3.3}$$

Theorem 3.1. *Let Assumption 1 hold. If f is bounded below, then Algorithm 2.1 terminates in finite iterations or satisfies*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.4}$$

Proof. The proof is similar to that of [9]. We prove the theorem by contradiction. Suppose it is not true, then there exists an integer $\hat{k}$ such that

$$\|g_k\| \ge \tau, \quad \forall k \ge \hat{k}. \tag{3.5}$$

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

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

Now we consider the following two cases.

Case (i). 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 Step 3 of Algorithm 2.1, we deduce

$$r_k < p_0, \quad \forall k \ge k_1.$$

Therefore by Step 4 of Algorithm 2.1 and (3.5), we have

$$\mu_k \to \infty, \quad \lambda_k \to \infty. \tag{3.6}$$

Since $x_{k+1} = x_k, \forall k \ge k_1$, we get from (2.9) and (3.6) that

$$\|d_k\| = \| - (G_k + \lambda_k I)^{-1} g_k\| \le \lambda_k^{-1}\|g_k\| \to 0. \tag{3.7}$$

From (2.11), we obtain

$$\begin{aligned} \|\hat{d}_k\| &= \| - (G_k + \lambda_k I)^{-1} g(y_k)\| \\ &\le \|(G_k + \lambda_k I)^{-1}(g(y_k) - g_k - G_k d_k)\| + \|(G_k + \lambda_k I)^{-1} g_k\| + \|(G_k + \lambda_k I)^{-1} G_k d_k\| \\ &\le L\lambda_k^{-1}\|d_k\|^2 + 2\|d_k\| \\ &\le C_1\|d_k\| \end{aligned} \tag{3.8}$$

for some positive constant C_1, where we use (3.3), (2.9) and $\|(G_k + \lambda_k I)^{-1} G_k\| \le 1$ in the second inequality, and the last inequality follows from (3.6) and (3.7).

It follows from (2.1) and (2.6) that

$$\begin{aligned} |\mathrm{Ared}_k - \mathrm{Pred}_k| &= \left| f(x_k) - f(x_k + d_k + \hat{d}_k) - \big(\varphi_{k,1}(0) - \varphi_{k,1}(d_k) + \varphi_{k,2}(0) - \varphi_{k,2}(\hat{d}_k)\big) \right| \\ &\le \left| f(y_k + \hat{d}_k) - f(y_k) - \frac{1}{2}\hat{d}_k^T G_k \hat{d}_k - g(y_k)^T \hat{d}_k \right| + \left| f(y_k) - f(x_k) - \frac{1}{2} d_k^T G_k d_k - g_k^T d_k \right| \\ &= o(\|d_k\|^2) + o(\|\hat{d}_k\|^2), \end{aligned} \tag{3.9}$$

where the last equality uses Taylor's formula, (3.2), (3.7) and (3.8).

Moreover, from (2.7), (3.5), (3.1) and (3.7), we have

$$\mathrm{Pred}_k \ge \frac{1}{2}\tau \min\left\{ \|d_k\|, \frac{\tau}{L} \right\} \ge \frac{1}{2}\tau\|d_k\| \tag{3.10}$$

for sufficiently large k.

Then the above two inequalities yield

$$\begin{aligned} |r_k - 1| &= \frac{|\mathrm{Ared}_k - \mathrm{Pred}_k|}{\mathrm{Pred}_k} \\ &= \frac{o(\|d_k\|^2) + o(\|\hat{d}_k\|^2)}{\|d_k\|} \to 0, \end{aligned} \tag{3.11}$$

which implies that $r_k \to 1$. Therefore from the parameter updating rule of Step 4 in Algorithm 2.1, there exists a positive constant C_2 such that

$$\mu_k \le C_2,$$

which contradicts to (3.6).

Case (ii). T is infinite. Then we have from (2.7) and (3.5) that

$$\begin{aligned}
\infty &> f(x_1) - \liminf_{k\to\infty} f(x_k) \ge \sum_{i=1}^{\infty} f(x_i) - f(x_{i+1}) \\
&= \sum_{k\in T} f(x_k) - f(x_{k+1}) \ge \sum_{k\in T} p_0 \mathrm{Pred}_k \\
&\ge \sum_{k\in T} p_0 \left(\frac{1}{2}\|g_k\| \min\left\{ \|d_k\|, \frac{\|g_k\|}{\|G_k\|} \right\} + \frac{1}{2}\|g(y_k)\| \min\left\{ \|\hat{d}_k\|, \frac{\|g(y_k)\|}{\|G_k\|} \right\} \right), \\
&\ge \sum_{k\in T} p_0 \frac{\tau}{2} \min\left\{ \|d_k\|, \frac{\tau}{L} \right\},
\end{aligned} \tag{3.12}$$

which implies that

$$\lim_{k\to\infty, k\in T} d_k = 0. \tag{3.13}$$

The above equality together with the updating rule of Step 4 in Algorithm 2.1 means

$$\lambda_k \to \infty. \tag{3.14}$$

Similar to (3.8), it follows from (3.13) and (3.14) that

$$\|\hat{d}_k\| \le C_3 \|d_k\|, \quad \forall k \in T \tag{3.15}$$

for some constant C_3. Then we have

$$\|s_k\| = \|d_k + \hat{d}_k\| \le (1 + C_3)\|d_k\|, \quad \forall k \in T. \tag{3.16}$$

This equality together with (3.12) yields

$$\sum_{k\in T} \|s_k\| < \infty, \tag{3.17}$$

which implies that

$$x_k \to x^*. \tag{3.18}$$

It follows from (2.9), (3.18), (3.14) and (3.8) that

$$d_k \to 0, \qquad \hat{d}_k \to 0. \tag{3.19}$$

Since $(G_k + \mu_k \|g_k\| I) d_k = -g_k$ from (2.9), we have from (3.5), (3.1) and (3.19) that

$$\mu_k \|d_k\| = \|g_k + G_k d_k\| \ge \|g_k\| - \|G_k\|\, \|d_k\| \ge \tau - L\|d_k\|,$$

which means

$$\mu_k \ge \frac{\tau}{\|d_k\|} - L \to \infty. \tag{3.20}$$

By the same analysis as (3.11) we know that

$$r_k \to 1,$$

which implies that there exists a constant C_4 such that

$$\mu_k \le C_4,$$

which leads to a contradiction to (3.20).

Based on the above analysis, we know that (3.4) holds. This finishes the proof. □

4. Local convergence

In this section, we suppose that $\{x_k\}$ converges to $x^* \in X$ and lies in some neighbourhood of x^*. We also give the following assumptions for local convergence analysis.

Assumption 2. (i) $\|g(x)\|$ provides a local error bound on some neighbourhood of x^*, i.e., there exist two positive constants c_1 and b_1 such that

$$\|g(x)\| \geq c_1 \mathrm{dist}(x, X), \quad \forall x \in N(x^*, b_1) = \{x | \ \|x - x^*\| \leq b_1\}. \tag{4.1}$$

(ii) The Hessian $G(x)$ is Lipschitz continuous on $N(x^*, b_1)$, that is, there exists a constant L such that

$$\|G(y) - G(x)\| \leq L\|y - x\|, \quad \forall x, y \in N(x^*, b_1). \tag{4.2}$$

It is clear that if $G(x)$ is nonsingular at a solution, then $\|g(x)\|$ provides a local error bound on its neighbourhood. However, the converse is not necessarily true [6,7], which shows that the local error bound condition is weaker than nonsingularity.

By Assumption 2, we have

$$\|g(y) - g(x)\| \leq L\|y - x\|, \quad \forall x, y \in N(x^*, b_1) \tag{4.3}$$

and

$$\|g(y) - g(x) - G(x)(y - x)\| \leq L\|y - x\|^2, \quad \forall x, y \in N(x^*, b_1). \tag{4.4}$$

In the later part of the paper, we denote $\bar{x} \in X$ which satisfies

$$\|\bar{x} - x\| = \mathrm{dist}(x, X) = \inf_{y \in X} \|y - x\|.$$

Since $G(x^*)$ is symmetric and positive semidefinite, there is an orthogonal matrix (U_1^*, U_2^*) such that

$$G(x^*) = (U_1^*, U_2^*) \begin{pmatrix} \Sigma_1^* & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1^{*T} \\ U_2^{*T} \end{pmatrix} = U_1^* \Sigma_1^* U_1^{*T}, \tag{4.5}$$

where Σ_1^* is a positive diagonal matrix.

Moreover, we can suppose that $G(x)$ has the following decomposition

$$G(x) = (U_1, U_2) \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} U_1{}^T \\ U_2{}^T \end{pmatrix} = U_1 \Sigma_1 U_1^T + U_2 \Sigma_2 U_2^T, \tag{4.6}$$

where $\mathrm{Rank}(\Sigma_1) = \mathrm{Rank}(\Sigma_1^*)$ and Σ_2 converges to zero as $x \to x^*$. In the following, for clearness, we also neglect the subscription k in the decomposition of $G(x_k)$, and still write $G(x_k)$ as same as (4.6).

In this section, we first prove the quadratic convergence of Algorithm 2.1, which implies that $\|x_k - x^*\|$ is equivalent to $\mathrm{dist}(x_k, X)$. Then we in turn show the cubic convergence of the proposed method.

4.1. Quadratic convergence

In this subsection, we first study the properties of $\|d_k\|$, $\|\hat{d}\|$ and $\|s_k\|$.

Lemma 4.1. *Let Assumption 2 hold. Then we have*

$$\begin{aligned} \|d_k\| &= O(\|\bar{x}_k - x_k\|), \\ \|\hat{d}_k\| &= O(\|\bar{x}_k - x_k\|), \\ \|s_k\| &= O(\|\bar{x}_k - x_k\|). \end{aligned} \tag{4.7}$$

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

$$\|\bar{x}_k - x_k\| = \mathrm{dist}(x_k, X) \leq \|x_k - x^*\| \to 0.$$

Moreover, the local error bound condition yields

$$\lambda_k = \mu_k \|g_k\| \geq m c_1 \mathrm{dist}(x_k, X) = m c_1 \|\bar{x}_k - x_k\|. \tag{4.8}$$

From (2.9), we get

$$\begin{aligned} \|d_k\| &= \|(G_k + \lambda_k I)^{-1} g_k\| \\ &\leq \|(G_k + \lambda_k I)^{-1}(g_k - g(\bar{x}_k) + G_k(\bar{x}_k - x_k))\| + \|(G_k + \lambda_k I)^{-1} G_k(\bar{x}_k - x_k)\| \\ &\leq L\lambda_k^{-1} \|\bar{x}_k - x_k\|^2 + \|\bar{x}_k - x_k\| \\ &= O(\|\bar{x}_k - x_k\|), \end{aligned} \tag{4.9}$$

where we use the fact $g(\bar{x}_k) = 0$ in the first inequality and (4.4) in the second inequality, and the last equality follows from (4.8).

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 (2.11), we obtain

$$\begin{aligned}
\|\hat{d}_k\| &= \|(G_k + \lambda_k I)^{-1} g(y_k)\| \\
&\le \|(G_k + \lambda_k I)^{-1}(g(y_k) - g_k - G_k d_k)\| + \|(G_k + \lambda_k I)^{-1} g_k\| + \|(G_k + \lambda_k I)^{-1} G_k d_k\| \\
&\le L\lambda_k^{-1}\|d_k\|^2 + 2\|d_k\| \\
&= O(\|\bar{x}_k - x_k\|),
\end{aligned}$$

where we use (4.4), (2.9), (4.8) and (4.9).

Thus we deduce from the above estimations of $\|d_k\|$ and $\|\hat{d}_k\|$ that

$$\|s_k\| = \|d_k + \hat{d}_k\| = O(\|\bar{x}_k - x_k\|). \quad \square$$

The following lemma shows that $\{\mu_k\}$ is bounded above.

Lemma 4.2. *Let Assumption 2 hold. There exists a constant c_2 such that*

$$\mu_k \le c_2.$$

Proof. From (2.3), (4.1) and (4.3), we have

$$\begin{aligned}
\varphi_{k,1}(0) - \varphi_{k,1}(d_k) &\ge \frac{1}{2}\|g_k\| \min\left\{\|d_k\|, \frac{\|g_k\|}{\|G_k\|}\right\} \\
&\ge \frac{1}{2} c_1 \|\bar{x}_k - x_k\| \min\left\{\|d_k\|, \frac{c_1}{L}\|\bar{x}_k - x_k\|\right\} \\
&\ge \bar{c}_2 \|\bar{x}_k - x_k\| \min\left\{\|d_k\|, \|\bar{x}_k - x_k\|\right\},
\end{aligned} \tag{4.10}$$

for some constant $\bar{c}_2$.

Then from (3.9), (2.7), (4.10) and Lemma 4.1, we get

$$\begin{aligned}
|r_k - 1| &= \left|\frac{\mathrm{Ared}_k - \mathrm{Pred}_k}{\mathrm{Pred}_k}\right| \\
&= \frac{o(\|d_k\|^2) + o(\|\hat{d}_k\|^2)}{\|\bar{x}_k - x_k\| \min\{\|d_k\|, \|\bar{x}_k - x_k\|\}} \to 0,
\end{aligned} \tag{4.11}$$

which implies that $r_k \to 1$. Hence we deduce from the updating rule of Step 4 in Algorithm 2.1 that there exists a constant c_2 such that $\mu_k \le c_2$. $\square$

Then we deduce that there exist constants c_3 and c_4 such that

$$c_3\|\bar{x}_k - x_k\| \le \lambda_k = \mu_k\|g_k\| = \mu_k\|g_k - g(\bar{x}_k)\| \le c_4\|\bar{x}_k - x_k\|, \tag{4.12}$$

which shows that $\|\bar{x}_k - x_k\|$ is equivalent to λ_k.

The following lemma means that $\{x_k\}$ converges to X quadratically.

Lemma 4.3. *Let Assumption 2 hold. Then we have*

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

Proof. From the local error bound condition, (4.4), (2.11) and (4.2), we have

$$\begin{aligned}
c_1\|\bar{x}_{k+1} - x_{k+1}\| &\le \|F(x_{k+1})\| \\
&= \|F(y_k + \hat{d}_k)\| \\
&\le \|F(y_k + \hat{d}_k) - F(y_k) - G(y_k)\hat{d}_k\| + \|F(y_k) + G(y_k)\hat{d}_k\| \\
&\le L\|\hat{d}_k\|^2 + \|F(y_k) + G_k\hat{d}_k\| + \|(G(y_k) - G_k)\hat{d}_k\| \\
&\le L\|\hat{d}_k\|^2 + \lambda_k\|\hat{d}_k\| + L\|d_k\|\,\|\hat{d}_k\| \\
&= O(\|\bar{x}_k - x_k\|^2),
\end{aligned} \tag{4.13}$$

where the last equality follows from Lemma 4.1 and (4.12). $\square$

The following theorem shows that $\{x_k\}$ converges to x^* quadratically, which is stronger than that of Lemma 4.3.

Theorem 4.1. *Let Assumption 2 hold. Then we have*

$$\|s_{k+1}\| = O(\|s_k\|^2), \qquad \|x_{k+1} - x^*\| = O(\|x_k - x^*\|^2).$$

Proof. From Lemma 4.3, it is clear for sufficiently large k that

$$\|\bar{x}_k - x_k\| \le \|\bar{x}_{k+1} - x_{k+1} + x_{k+1} - x_k\| \le \|\bar{x}_{k+1} - x_{k+1}\| + \|s_k\| \le 2\|s_k\|. \tag{4.14}$$

This inequality together with Lemmas 4.1 and 4.3 yields

$$\|s_{k+1}\| = O(\|s_k\|^2), \tag{4.15}$$

which implies that

$$\|x_{k+1} - x^*\| = O(\|x_k - x^*\|^2). \quad \square$$

4.2. Cubic convergence

To obtain faster convergence of the proposed method, we need to estimate $\|\hat{d}_k\|$ more accurately. The following lemma shows that $\|s_k\|$ is equivalent to $\|x_k - x^*\|$ if x_k converges to x^* superlinearly, where $s_k = x_{k+1} - x_k$.

Lemma 4.4 ([2, Theorem 1.5.2])**.** *If the sequence $\{x_k\}$ converges superlinearly to x^*, then*

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

Therefore we have from Lemma 4.4, Theorem 4.1 and Lemma 4.1 that there exist two positive constants c_5 and c_6 such that

$$\|x_k - x^*\| \le c_5\|s_k\| \le c_6\|\bar{x}_k - x_k\| \le c_6\|x_k - x^*\|, \tag{4.16}$$

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

By the theory of matrix perturbation [14] and (4.2), we have

$$\|\Sigma_1 - \Sigma_1^*\| + \|\Sigma_2\| \le \|G_k - G(x^*)\| \le L\|x_k - x^*\|.$$

This inequality together with (4.16) yields

$$\|\Sigma_1 - \Sigma_1^*\| \le L\|\bar{x}_k - x_k\|, \qquad \|\Sigma_2\| \le L\|\bar{x}_k - x_k\|. \tag{4.17}$$

Lemma 4.5. *Let Assumption 2 hold. Then we have*

$$\begin{aligned} &\|g(y_k)\| = O(\|\bar{x}_k - x_k\|^2), \\ &\|U_2U_2^T g(y_k)\| = O(\|\bar{x}_k - x_k\|^3). \end{aligned} \tag{4.18}$$

Proof. From (2.9), (4.12) and Lemma 4.1, we have

$$\|g_k + G_k d_k\| = \lambda_k\|d_k\| = O(\|\bar{x}_k - x_k\|^2). \tag{4.19}$$

Similarly, we know

$$\|g(y_k) + G_k\hat{d}_k\| = \lambda_k\|\hat{d}_k\| = O(\|\bar{x}_k - x_k\|^2). \tag{4.20}$$

Then we get from (4.4), (4.19) and Lemma 4.1 that

$$\begin{aligned} \|g(y_k)\| &= \|g(x_k + d_k) - g_k - G_k d_k\| + \|g_k + G_k d_k\| \\ &= O(\|\bar{x}_k - x_k\|^2). \end{aligned} \tag{4.21}$$

From the local error bound condition and (4.21), we have

$$\|\bar{y}_k - y_k\| \le c_1^{-1}\|g(y_k)\| = O(\|\bar{x}_k - x_k\|^2). \tag{4.22}$$

Set $\tilde{G}_k = U_1\Sigma_1U_1^T$ and $\tilde{d}_k = -\tilde{G}_k^+ g(y_k)$, then $\tilde{d}_k$ is the least square solution of

$$\min \|g(y_k) + \tilde{G}_k d\|.$$

Therefore we have

$$\begin{aligned}\|U_2U_2^Tg(y_k)\| &= \|g(y_k)+\tilde{G}_k\tilde{d}_k\| \le \|g(y_k)+\tilde{G}_k(\bar{y}_k-y_k)\| \\ &\le \|g(y_k)+G(y_k)(\bar{y}_k-y_k)\| + \|(G(y_k)-G_k)(\bar{y}_k-y_k) + \|(G_k-\tilde{G}_k)(\bar{y}_k-y_k)\ \| \\ &\le L\|\bar{y}_k-y_k\|^2 + L\|d_k\|\ \|\bar{y}_k-y_k\| + \|\Sigma_2\|\ \|\bar{y}_k-y_k\| \\ &= O(\|\bar{x}_k-x_k\|^3),\end{aligned}$$

where we use the fact $g(\bar{y}_k)=0$, (4.4) and (4.2) in the third inequality, and the last equality follows from (4.22), (4.17) and Lemma 4.1. □

Lemma 4.6. *Let Assumption 2 hold. Then we have*

$$\|\hat{d}_k\| = O(\|\bar{x}_k-x_k\|^2). \tag{4.23}$$

Proof. From (2.11), we have

$$\begin{aligned}\hat{d}_k &= -(G_k+\lambda_kI)^{-1}g(y_k) \\ &= -U_1(\Sigma_1+\lambda_kI)^{-1}U_1^Tg(y_k) - U_2(\Sigma_2+\lambda_kI)^{-1}U_2^Tg(y_k).\end{aligned} \tag{4.24}$$

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 c_7 such that

$$\|\Sigma_1^{-1}\| \le c_7. \tag{4.25}$$

Then from (4.24), (4.25), (4.12) and Lemma 4.5, we obtain

$$\begin{aligned}\|\hat{d}_k\| &\le \|\Sigma_1^{-1}\|\ \|U_1U_1^Tg(y_k)\| + \lambda_k^{-1}\|U_2U_2^Tg(y_k)\| \\ &\le c_7\|g(y_k)\| + \lambda_k^{-1}\|U_2U_2^Tg(y_k)\| \\ &= O(\|\bar{x}_k-x_k\|^2). \quad \square\end{aligned}$$

Now from the local error bound condition again, we get

$$\begin{aligned}c_1\|\bar{x}_{k+1}-x_{k+1}\| &\le \|g(x_{k+1})\| = \|g(y_k+\hat{d}_k)\| \\ &\le \|g(y_k+\hat{d}_k)-g(y_k)-G(y_k)\hat{d}_k\| + \|G(y_k)\hat{d}_k+g(y_k)\| \\ &\le L\|\hat{d}_k\|^2 + \|(G(y_k)-G_k)\hat{d}_k\| + \|G_k\hat{d}_k+g(y_k)\| \\ &\le L\|\hat{d}_k\|^2 + L\|d_k\|\ \|\hat{d}_k\| + \lambda_k\|\hat{d}_k\| \\ &= O(\|\bar{x}_k-x_k\|^3),\end{aligned} \tag{4.26}$$

where we use (4.4), (4.2) and (2.11) in the fourth inequality, and the last equality follows from Lemmas 4.6 and 4.1 and (4.12).

From Lemma 4.1, (4.14) and (4.26), we have

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

which implies that $\{x_k\}$ converges to x^* cubically. We summarize this main result as follows.

Theorem 4.2. *Let Assumption 2 hold. Then the sequence $\{x_k\}$ generated by Algorithm 2.1 converges cubically, that is, $\|x_{k+1}-x^*\| = O(\|x_k-x^*\|^3)$.*

Acknowledgements

This work was supported by the NSF (10901026) of China and the Key Project of the Scientific Research Fund of the Hunan Provincial Education Department.

References

[1] C.T. Kelley, Iterative Methods for Optimization, in: Frontiers in Applied Mathematics, vol. 18, SIAM, Philadelphia, 1999.
[2] W. Sun, Y. Yuan, Optimization Theory and Methods, Springer Science and Business Media, LLC, New York, 2006.
[3] W. Zhou, X. Chen, Global convergence of a new hybrid Gauss–Newton structured BFGS methods for nonlinear least squares problems, SIAM J. Optim. 20 (2010) 2422–2441.
[4] W. Zhou, D. Li, A globally convergent BFGS method for nonlinear monotone equations without any merit functions, Math. Comp. 77 (2008) 2231–2240.
[5] J. Fan, Y. Yuan, On the quadratic convergence of the Levenberg–Marquardt method without nonsingularity assumption, Computing 74 (2005) 23–39.

[6] D. Li, M. Fukushima, L. Qi, N. Yamashita, Regularized Newton methods for convex minimization problems with singular solutions, Comput. Optim. Appl. 28 (2004) 131–147.
[7] N. Yamashita, M. Fukushima, On the rate of convergence of the Levenberg–Marquardt method, Computing 15 (Suppl.) (2001) 237–249.
[8] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM, Philadelphia, 2005.
[9] J. Fan, The modified Levenberg–Marquardt method for nonlinear equations with cubic convergence, Math. Comp. 81 (2012) 447–466.
[10] Y. Li, D. Li, Truncated regularized Newton method for convex minimization, Comput. Optim. Appl. 43 (2009) 119–131.
[11] G. Zhou, L. Qi, On the convergence of an inexact Newton-type method, Oper. Res. Lett. 34 (2006) 647–652.
[12] G. Zhou, K.C. Toh, Superlinear convergence of a Newton-type algorithm for monotone equations, J. Optim. Theory Appl. 125 (2005) 205–221.
[13] 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.
[14] G.W. Stewart, J.G. Sun, Matrix Perturbation Theory, Academic Press, San Diego, CA, 1990.