On the convergence of the modified Levenberg-Marquardt method with a nonmonotone second order Armijo type line search

Weijun Zhou*

Abstract. Recently, Fan [4, Math. Comput., 81 (2012), pp. 447-466] proposed a modified Levenberg-Marquardt (MLM) method for nonlinear equations. Using a trust region technique, global and cubic convergence of the MLM method is proved [4] under the local error bound condition, which is weaker than nonsingularity. The purpose of the paper is to investigate the convergence properties of the MLM method with a line search technique. Since the search direction of the MLM method may be not a descent direction, standard line searches can not be used directly. In this paper, we propose a nonmonote second order Armijo line search which guarantees the global convergence of the MLM method. Moreover, we prove that the unit step will be always accepted finally. Then cubic convergence of the MLM method is preserved under the local error bound condition.
Keywords. Nonlinear equations, line search, global convergence, cubic convergence.
AMS subject classification. 65K05, 90C30.

1 Introduction

Let $F : R^n \to R^n$ be a continuously differentiable mapping. Consider the system of nonlinear equations

$$F(x) = 0, \tag{1.1}$$

which is one of the cornerstones of computation mathematics. 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. Many efficient solution techniques like the Newton method, quasi-Newton methods, the Gauss-Newton method, the Levenberg-Marquardt method, etc. are available for this problem such as [1, 3, 5, 8, 10, 11, 12, 14, 15, 16, 17].

In this paper, we focus on the Levenberg-Marquardt (LM) method, which computes the search direction by

$$d_k^{LM} = -(J_k^T J_k + \lambda_k I)^{-1} J_k^T F_k,$$

where λ_k is a nonnegative regularized parameter, $F_k = F(x_k)$ and $J_k = F'(x_k)$ is the Jacobian of F at x_k. It is well-known that the LM method has quadratic convergence as the Newton method if the Jacobian is Lipschitz continuous and nonsingular at the solution.

However, the condition on the nonsingularity of the Jacobian is very strong. Recently, under the local error bound condition which is weaker than nonsingularity [15], Fan [4] proposed a modified Levenberg-Marquardt (MLM) method with cubic convergence. At each iteration, the MLM method first obtains d_k^{LM} by solving the following linear equations

$$(J_k^T J_k + \lambda_k I)d = -J_k^T F_k \quad \text{with} \quad \lambda_k = \mu_k \|F_k\|^{\delta}, \delta \in [1, 2],$$

where $\mu_k > 0$ is updated from step by step using a trust region technique, then solves the linear equations

$$(J_k^T J_k + \lambda_k I)d = -J_k^T F(y_k) \quad \text{with} \quad y_k = x_k + d_k^{LM}$$

*Department of Mathematics, Changsha University of Science and Technology, Changsha 410004, China(weijunzhou@126.com). This author's work was supported by the NSF foundation (10901026) of China and The Open Fund Project of Key Research Institute of Philosophies and Social Sciences in Hunan Universities.

to get the approximate LM step d_k^{MLM} and set the search direction $d_k = d_k^{LM} + d_k^{MLM}$. It is clear that d_k may be not necessarily a descent direction of the merit function $\|F(x)\|^2$. Fan [4] obtained the global convergence of the MLM method using a trust region technique.

However, it is not easy to prove the global convergence of the MLM method when using line search since d_k is no longer a descent direction. And hence standard line search techniques can not be used directly in this case.

The purpose of the paper is to investigate this problem, that is, with some line search, whether the global and cubic convergence of the MLM method can be preserved as the trust region case under the local error bound condition.

First let us simply recall some nonmonotone line search techniques. The best known nomonotone line search was proposed by Grippo, Lampariello and Lucidi [7] for optimization, which can be written as follows:

$$f(x_k + \alpha_k d_k) \le \max_{0 \le j \le M-1} f(x_{k-j}) + \gamma \nabla f(x_k)^T d_k,$$

where M is a nonnegative integer, $\gamma \in (0,1)$ and f is a merit function such that $f(x) = 0$ if and only if $\|F(x)\| = 0$. However this line search is only suitable for descent methods which satisfy $\nabla f(x_k)^T d_k < 0$.

Li and Fukushima [10] presented a nonmonotone line search for nonlinear equations, that is,

$$\|F(x_k + \alpha_k d_k)\|^2 - \|F(x_k)\|^2 \le -\sigma_1 \|\alpha_k F(x_k)\|^2 - \sigma_2 \|\alpha_k d_k\|^2 + \epsilon_k \|F(x_k)\|^2, \tag{1.2}$$

where σ_1, σ_2 are positive constants and the positive sequence $\{\epsilon_k\}$ satisfies (1.8). This line search can avoid the necessity of descent directions to ensure that each iteration is well defined. However it is not suitable for the MLM method since the direction of this method contains two parts. Hence we need to modify this line search for the MLM method.

Note that Goldfarb [6, 14] proposed a second order Armijo step rule for the negative curvature direction method for solving optimization, which is given by

$$f(x_k + \alpha_k s_k + \alpha^2 d_k) \le \gamma\big(\alpha_k s_k^T \nabla f(x_k) + \frac{1}{2}\alpha_k^4 d_k^T \nabla^2 f(x_k) d_k\big), \tag{1.3}$$

where $\gamma \in (0,1)$ and (s_k, d_k) is a descent pair(please see [6, 14]) of the objective function f at x_k.

Motivated by (1.2) and (1.3), in this paper, we propose a new nonmonotone second order Armijo type line search (1.7) below. Now it is convenient for us to present the complete algorithm with this new line search as follows.

Algorithm 1.1 (The MLM method with line search).

> **Step 1.** Choose a starting point $x_0 \in R^n$ and several constants $\mu > 0$, $\sigma_1, \sigma_2, \sigma_3 > 0$ and $r, \rho \in (0,1)$. Let $k := 0$.
>
> **Step 2.** If $\|J_k^T F_k\| = 0$, then stop. Otherwise compute d_k by solving the following linear equations
>
> $$(J_k^T J_k + \lambda_k I)d = -J_k^T F_k \quad \text{with} \quad \lambda_k = \mu \|F_k\|. \tag{1.4}$$
>
> Then solve the following linear equations to obtain $\hat{d}_k$:
>
> $$(J_k^T J_k + \lambda_k I)d = -J_k^T F(y_k) \quad \text{with} \quad y_k = x_k + d_k. \tag{1.5}$$

Step 3. If

$$\|F(x_k + d_k + \hat{d}_k)\| \le \rho\|F_k\|, \tag{1.6}$$

then take $\alpha_k = 1$ and go to Step 5. Otherwise go to Step 4.

Step 4. Compute $\alpha_k = \max\{1, r^1, r^2, \cdots\}$ with $\alpha = r^i$ satisfying

$$\begin{aligned} & \|F(x_k + \alpha d_k + \alpha^2 \hat{d}_k)\|^2 - \|F_k\|^2 \\ \le \quad & -\sigma_1\alpha^2\|d_k\|^2 - \sigma_2\alpha^2\|\hat{d}\|^2 - \sigma_3\alpha^2\|F_k\|^2 + \epsilon_k\|F_k\|^2, \end{aligned} \tag{1.7}$$

where $\{\epsilon_k\}$ is a given positive sequence such that

$$\sum_{k=0}^{\infty} \epsilon_k < \infty. \tag{1.8}$$

Step 5. Set $x_{k+1} = x_k + \alpha_k d_k + \alpha_k^2 \hat{d}_k$. Let $k := k + 1$ and go to Step 2.

Remark: (i) It is clear that as $\alpha \to 0^+$, the left-hand side of (1.7) goes to zero, while the right-hand side tends to the positive $\epsilon_k\|F_k\|^2$. Thus (1.7) is satisfied for all sufficiently small $\alpha > 0$. Then the algorithm is well defined.

(ii) The computation cost of the MLM method is almost as same as that of the standard LM method since (1.5) only involves $F(y_k)$ and can use the available decomposition of $J_k^T J_k + \lambda_k I$ after solving (1.4).

(iii) In [4], the parameter $\lambda_k = \mu_k\|F_k\|^\delta$ with $\delta \in [1, 2]$. In Algorithm 1.1, we only set $\lambda_k = \mu\|F_k\|$ which is also suggested by Kelley in his book [9].

The paper is organized as follows. In Section 2, we show the global convergence of Algorithm 1.1 under suitable conditions. In Section 3, we prove that $\alpha_k \equiv 1$ for sufficiently large k. And hence the cubic convergence of Algorithm 1.1 is still preserved under the local error bound condition.

2 Global convergence

Define the level set

$$\Omega = \{x | \ \|F(x)\| \le e^{\frac{\epsilon}{2}}\|F_0\|\}, \tag{2.1}$$

where ϵ is a positive constant such that

$$\sum_{k=0}^{\infty} \epsilon_k \le \epsilon < \infty. \tag{2.2}$$

Lemma 2.1. [2, Lemma 3.3] *Let $\{a_k\}$ and $\{r_k\}$ be positive sequences satisfying $a_{k+1} \le (1 + r_k)a_k + r_k$ and $\sum_{k=0}^{\infty} r_k < \infty$. Then $\{a_k\}$ converges.*

Then we have the following lemma whose proof is similar to that of Lemma 2.1 in [10], however for completeness, we give the proof here.

Lemma 2.2. *Let the sequence $\{x_k\}$ be generated by Algorithm 1.1, then the sequence $\{\|F_k\|\}$ converges and $x_k \in \Omega$ for all $k \ge 0$.*

Proof. From (1.6) and (1.7), we have

$$\|F_{k+1}\|^2 \le (1 + \epsilon_k)\|F_k\|^2,$$

which together with (2.2) and Lemma 2.1 implies that $\{\|F_k\|^2\}$ converges. Hence $\{\|F_k\|\}$ also converges.

Moreover, from (2.2), we deduce that

$$\begin{aligned}\|F_{k+1}\| &\leq (1+\epsilon_k)^{\frac{1}{2}}\|F_k\| \leq \cdots \leq \Pi_{i=0}^{k}(1+\epsilon_i)^{\frac{1}{2}}\|F_0\| \\ &\leq \Big(\sum_{i=0}^{k}\frac{1}{k+1}(1+\epsilon_i)\Big)^{\frac{k+1}{2}}\|F_0\| \\ &\leq \Big(1+\frac{\epsilon}{k+1}\Big)^{\frac{k+1}{2}}\|F_0\| \\ &\leq e^{\frac{\epsilon}{2}}\|F_0\|,\end{aligned}$$

which means $x_k \in \Omega$ for all k. The proof is completed. □

It is clear that Lemma 2.2 implies the sequence $\{\|F_k\|\}$ is bounded, that is, there exists a constant $M > 0$ such that

$$\|F_k\| \leq M, \quad \forall k \geq 0. \tag{2.3}$$

In this section, we make the following assumptions to study the global convergence of Algorithm 1.1.

Assumption 2.1 There exists a neighbourhood Ω_1 of Ω such that $F(x)$ and its Jacobian $J(x)$ are Lipschitz continuous, that is, there exists a positive constant L such that

$$\|J(x)-J(y)\| \leq L\|x-y\|, \quad \forall x,y \in \Omega_1, \tag{2.4}$$

and

$$\|F(x)-F(y)\| \leq L\|x-y\|, \quad \forall x,y \in \Omega_1. \tag{2.5}$$

From (2.5), we have

$$\|J(x)\| \leq L, \quad \forall x \in \Omega_1. \tag{2.6}$$

In fact, for $x \in \Omega_1$, $h \in R^n$ and $t > 0$, we have

$$\begin{aligned}\|J(x)h\| &= \Big\|\frac{F(x+th)-F(x)}{t} - \Big(\frac{F(x+th)-F(x)}{t} - J(x)h\Big)\Big\| \\ &\leq \Big\|\frac{F(x+th)-F(x)}{t}\Big\| + \Big\|\frac{F(x+th)-F(x)}{t} - J(x)h\Big\| \\ &\leq L\|h\| + \Big\|\frac{F(x+th)-F(x)}{t} - J(x)h\Big\|,\end{aligned}$$

where the second inequality uses (2.5). Let $t \to 0^+$, by the differentiability of F, we have

$$\|J(x)h\| \leq L\|h\|,$$

which implies (2.6).

Now we give the following global convergence result for Algorithm 1.1.

Theorem 2.1. *Let Assumption 2.1 hold. Then Algorithm 1.1 terminates in finite iterations or satisfies*

$$\liminf_{k\to\infty} \|J_k^T F_k\| = 0. \tag{2.7}$$

Proof. We prove the theorem by contradiction. Suppose it is not true, then there exists an integer $\hat{k}$ such that

$$\|J_k^T F_k\| \geq \tau, \quad \forall k \geq \hat{k}, \tag{2.8}$$

which implies that

$$\|F_k\| \geq \tau_1 \tag{2.9}$$

holds for sufficiently large k with some positive constant τ_1.

If (1.6) holds for infinite k, then $\|F_k\| \to 0$, which is a contradiction to (2.9). In fact, denote the index sets

$$H_j = \{k \leq j |\ (1.6) \text{ holds}\}, \quad G_j = \{0, 1, \cdots, j\} \backslash H_j, \quad j = 1, 2, \cdots.$$

If (1.6) holds for infinite k, then as $j \to \infty$,

$$|H_j| \to \infty,$$

where $|H_j|$ is the number of the set H_j. From (1.6) and (1.7), we have

$$\begin{aligned}
\|F_{k+1}\| &\leq \prod_{i \in G_k} (1 + \epsilon_i)^{1/2} \prod_{i \in H_k} \rho \|F_0\| \\
&= \prod_{i \in G_k} (1 + \epsilon_i)^{1/2} \rho^{|H_k|} \|F_0\| \\
&\leq e^{\epsilon/2} \rho^{|H_k|} \|F_0\| \to 0, \quad \text{as } k \to \infty.
\end{aligned}$$

From now on we assume (1.6) holds only for finite k. Therefore we obtain from (1.7) that

$$\sum_{k=0}^{\infty} \alpha_k^2 \|d_k\|^2 < \infty, \quad \sum_{k=0}^{\infty} \alpha_k^2 \|\hat{d}_k\|^2 < \infty, \quad \sum_{k=0}^{\infty} \alpha_k^2 \|F_k\|^2 < \infty,$$

which imply

$$\lim_{k \to \infty} \alpha_k \|d_k\| = 0, \quad \lim_{k \to \infty} \alpha_k \|\hat{d}_k\| = 0, \quad \lim_{k \to \infty} \alpha_k \|F_k\| = 0.$$

These equalities together with (2.9) yield

$$\lim_{k \to \infty} \alpha_k = 0. \tag{2.10}$$

Set $\bar{\alpha}_k = \alpha_k / r$. Then from the line search (1.7), we have

$$\begin{aligned}
&\|F(x_k + \bar{\alpha}_k d_k + \bar{\alpha}_k^2 \hat{d}_k\|^2 - \|F_k\|^2 \\
\geq\ & -\sigma_1 \bar{\alpha}_k^2 \|d_k\|^2 - \sigma_2 \bar{\alpha}_k^2 \|\hat{d}_k\|^2 - \sigma_3 \bar{\alpha}_k^2 \|F_k\|^2 + \epsilon_k \|F_k\|^2 \\
\geq\ & -\sigma_1 \bar{\alpha}_k^2 \|d_k\|^2 - \sigma_2 \bar{\alpha}_k^2 \|\hat{d}_k\|^2 - \sigma_3 \bar{\alpha}_k^2 \|F_k\|^2,
\end{aligned}$$

which means that

$$\begin{aligned}
&\bar{\alpha}_k^2 \big(\sigma_1 \|d_k\|^2 + \sigma_2 \|\hat{d}_k\|^2 + \sigma_3 \|F_k\|^2\big) \\
\geq\ & -\big(\|F(x_k + \bar{\alpha}_k d_k + \bar{\alpha}_k^2 \hat{d}_k)\|^2 - \|F_k\|^2\big) \\
=\ & -\Big(2F_k^T (F(x_k + \bar{\alpha}_k d_k + \bar{\alpha}_k^2 \hat{d}_k) - F_k) + \|F(x_k + \bar{\alpha}_k d_k + \bar{\alpha}_k^2 \hat{d}_k) - F_k\|^2\Big) \\
\geq\ & -2F_k^T (F(x_k + \bar{\alpha}_k d_k + \bar{\alpha}_k^2 \hat{d}_k) - F_k) - C_1 \bar{\alpha}_k^2 (\|d_k\|^2 + \|\hat{d}_k\|^2),
\end{aligned} \tag{2.11}$$

for some positive constant C_1, where the last inequality uses (2.5) and the fact $\bar{\alpha}_k \leq \frac{1}{r}$.

Now we estimate the term $F_k^T(F(x_k + \bar{\alpha}_k d_k + \bar{\alpha}_k^2 \hat{d}_k) - F_k)$. Note that

$$\begin{aligned}
& F_k^T\big(F(x_k + \bar{\alpha}_k d_k + \bar{\alpha}_k^2 \hat{d}_k) - F_k\big) \\
= \;& F_k^T\big(F(x_k + \bar{\alpha}_k d_k + \bar{\alpha}_k^2 \hat{d}_k) - F(x_k + \bar{\alpha}_k d_k)\big) + F_k^T\big(F(x_k + \bar{\alpha}_k d_k) - F_k\big) \\
\leq \;& LM\bar{\alpha}_k^2\|\hat{d}_k\| + F_k^T J_k \bar{\alpha}_k d_k + F_k^T \int_0^1 \big(J(x_k + t\bar{\alpha}_k d_k) - J_k\big)\bar{\alpha}_k d_k dt \\
\leq \;& 2LM\bar{\alpha}_k^2\|\hat{d}_k\| - \bar{\alpha}_k d_k^T(J_k^T J_k + \lambda_k I)d_k, \qquad (2.12)
\end{aligned}$$

where the first inequality uses (2.5) and (2.3), the last inequality uses (1.4) and (2.4).

Then from (2.11)-(2.12), we get that there exists a positive constant C_2 such that

$$\bar{\alpha}_k \geq \frac{d_k^T(J_k^T J_k + \lambda_k I)d_k}{C_2(\|d_k\|^2 + \|\hat{d}_k\|^2 + \|F_k\|^2 + \|\hat{d}_k\|)} \geq \frac{\lambda_k d_k^T d_k}{C_2(\|d_k\|^2 + \|\hat{d}_k\|^2 + \|F_k\|^2 + \|\hat{d}_k\|)}. \quad (2.13)$$

Let the SVD of J_k be

$$J_k = U\Sigma V^T,$$

where U, V are two orthogonal matrixes, and Σ is a diagonal matrix with nonnegative $\sigma_i \geq 0$, $i = 1, \cdots, n$. Then

$$\begin{aligned}
\|(J_k^T J_k + \lambda_k I)^{-1}\| &= \|V(\Sigma^2 + \lambda_k I)^{-1} V^T\| \\
&= \|(\Sigma^2 + \lambda_k I)^{-1}\| \\
&= \max_{i \in \{1,2\cdots,n\}} (\sigma_i^2 + \lambda_k)^{-1} \\
&\leq \lambda_k^{-1}.
\end{aligned}$$

From (1.4), (2.5), (2.6) and the above inequality, we have

$$\|d_k\| = \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T F_k\| \leq \|(J_k^T J_k + \lambda_k I)^{-1}\|\|J_k\|\|F_k\| \leq L\lambda_k^{-1}\|F_k\| = \frac{L}{\mu}. \quad (2.14)$$

Similarly, from (1.5), (2.9) and (2.5), we obtain

$$\begin{aligned}
\|\hat{d}_k\| &= \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T F(y_k)\| \\
&\leq \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T (F(y_k) - F_k)\| + \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T F_k\| \\
&\leq L^2\lambda_k^{-1}\|d_k\| + \|d_k\| \leq \Big(1 + \frac{L^2}{\mu\tau_1}\Big)\|d_k\|. \qquad (2.15)
\end{aligned}$$

If $\liminf_{k\to\infty} \|d_k\| = 0$, then we have from (1.4) and (2.3) that

$$\liminf_{k\to\infty} \|J_k^T F_k\| = \liminf_{k\to\infty} \|(J_k^T J_k + \lambda_k I)d_k\| = 0,$$

which contradics to (2.8). Hence there exists a constant $\tau_2 > 0$ such that

$$\liminf_{k\to\infty} \|d_k\| \geq \tau_2,$$

which together with (2.13)-(2.15), (2.3) and (2.9) implies that $\{\alpha_k\}$ is bounded away from zero. This leads to a contradiction to (2.10). The proof is then finished. □

3 Cubic convergence

In this section, we assume that $x_k \to x^* \in X$ and the sequence $\{x_k\}$ lies in some neighbourhood of x^*. The key to the local convergence is to show that the unit step will be taken for all sufficiently large k. We give the following assumptions as same as those of [4] for the local convergence analysis.

Assumption 3.1 (i) $\|F(x)\|$ provides a local error bound on some neighbourhood of x^*, i.e., there exist two positive constant c_1 and b_1 such that

$$\|F(x)\| \geq c_1 \text{dist}(x, X), \quad \forall x \in N(x^*, b_1) = \{x | \; \|x - x^*\| \leq b_1\}. \tag{3.1}$$

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

$$\|J(y) - J(x)\| \leq L\|y - x\|, \quad \forall x, y \in N(x^*, b_1). \tag{3.2}$$

It is clear that if $J(x)$ is nonsingular at a solution, then $\|F(x)\|$ provides a local error bound on its neighbourhood. However, the converse is not necessarily true [15], which shows that the local error bound condition is weaker than nonsingularity.

By Assumption 3.1, we have

$$\|F(y) - F(x)\| \leq L\|y - x\|, \quad \forall x, y \in N(x^*, b_1) \tag{3.3}$$

and

$$\|F(y) - F(x) - J(x)(y - x)\| \leq L\|y - x\|^2, \quad \forall x, y \in N(x^*, b_1). \tag{3.4}$$

In the following, we denote $\bar{x} \in X$ such that

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

From the local error bound condition and (3.3), we have

$$c_1 \mu \|\bar{x}_k - x_k\| \leq \lambda_k = \mu \|F_k\| \leq L\mu \|\bar{x}_k - x_k\|. \tag{3.5}$$

Now suppose the SVD of $J(x^*)$ is

$$J(x^*) = (U_1^*, U_2^*) \begin{pmatrix} \Sigma_1^* & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^{*T} \\ V_2^{*T} \end{pmatrix} = U_1^* \Sigma_1^* V_1^{*T}, \tag{3.6}$$

where (U_1^*, U_2^*) and (V_1^*, V_2^*) are two orthogonal matrixes, and Σ_1^* is a diagonal matrix with positive diagonals. Correspondingly, we can suppose that the SVD of $J(x)$ has the following form

$$J(x) = (U_1, U_2) \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T, \tag{3.7}$$

where $\text{Rank}(\Sigma_1) = \text{Rank}(\Sigma_1^*)$ 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 $J(x_k)$, and still write $J(x_k)$ as same as (3.7).

By the matrix perturbation theory [13] and (3.2), we have

$$\|\Sigma_1 - \Sigma_1^*\| + \|\Sigma_2\| \leq \|J_k - J(x^*)\| \leq L\|x_k - x^*\|,$$

which implies

$$\|\Sigma_1 - \Sigma_1^*\| \leq L\|x_k - x^*\|, \quad \|\Sigma_2\| \leq L\|x_k - x^*\|. \tag{3.8}$$

Note that $x_k \to x^*$, then there exist a constant C_3 such that

$$\|\Sigma_1\| \le C_3, \quad \|\Sigma_1^{-1}\| \le C_3. \tag{3.9}$$

Then we deduce

$$\begin{aligned}
& \|(J_k^T J_k + \lambda_k I)^{-1} J_k\| \\
= & \left\| (V_1, V_2) \begin{pmatrix} (\Sigma_1^2 + \lambda_k I)^{-1}\Sigma_1 & \\ & (\Sigma_2^2 + \lambda_k I)^{-1}\Sigma_2 \end{pmatrix} \begin{pmatrix} U_1{}^T \\ U_2{}^T \end{pmatrix} \right\| \\
\le & \left\| \begin{pmatrix} (\Sigma_1^2 + \lambda_k I)^{-1}\Sigma_1 & \\ & (\Sigma_2^2 + \lambda_k I)^{-1}\Sigma_2 \end{pmatrix} \right\| \\
\le & \left\| \begin{pmatrix} \Sigma_1^{-1} & \\ & \lambda_k^{-1}\Sigma_2 \end{pmatrix} \right\| \\
\le & \; C_3 + \frac{L\|x_k - x^*\|}{c_1\mu\|\bar{x}_k - x_k\|},
\end{aligned} \tag{3.10}$$

where the last inequality follows from (3.9), (3.8) and (3.5). Then we have the following lemma.

Lemma 3.3. *Let Assumption 3.1 hold, then we have*

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

Proof. From (1.4), it is easy to see that d_k is also the minimizer of the following convex optimization problem:

$$\min_{d \in R^n} \varphi_{k,1}(d) = \|F_k + J_k d\|^2 + \lambda_k \|d\|^2. \tag{3.11}$$

Thus from (3.11), $F(\bar{x}_k) = 0$, (3.4) and (3.5), we get

$$\begin{aligned}
\|d_k\|^2 &\le \frac{\varphi_{k,1}(d_k)}{\lambda_k} \le \frac{\varphi_{k,1}(\bar{x}_k - x_k)}{\lambda_k} \\
&= \frac{\|F_k + J_k(\bar{x}_k - x_k)\|^2}{\lambda_k} + \|\bar{x}_k - x_k\|^2 \\
&\le C_4\|\bar{x}_k - x_k\|^3 + \|\bar{x}_k - x_k\|^2
\end{aligned}$$

for some positive constant C_4, which implies that

$$\|d_k\| = O(\|\bar{x}_k - x_k\|). \tag{3.12}$$

From (1.5), (3.4), (1.4), (3.10) and (3.12), we obtain

$$\begin{aligned}
\|\hat{d}_k\| &= \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T F(y_k)\| \\
&\le \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T (F(y_k) - F_k - J_k d_k)\| \\
&\quad + \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T F_k\| + \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T J_k d_k\| \\
&\le L\|d_k\|^2 \|(J_k^T J_k + \lambda_k I)^{-1} J_k\| + 2\|d_k\| \\
&\le C_5\|d_k\|\|x_k - x^*\| + C_5\|d_k\| \\
&= O(\|\bar{x}_k - x_k\|)
\end{aligned}$$

for some positive constant C_5. This finishes the proof. □

Lemma 3.4. *Let Assumption 3.1 hold, then we have*

$$\begin{aligned} \|U_2U_2^TF_k\| &= o(\|\bar{x}_k - x_k\|), \\ \|F(y_k)\| &= o(\|\bar{x}_k - x_k\|). \end{aligned}$$

Proof. Let $\tilde{J}_k = U_1\Sigma_1V_1^T$ and $\tilde{d}_k = -\tilde{J}_k^+F_k$, where $\tilde{J}_k^+$ is the Moore-Penrose generalized inverse of $\tilde{J}_k$. Then $\tilde{d}_k$ is the least squares solution of $\min \|F_k + \tilde{J}_kd\|$. From (3.4) and (3.8), we have

$$\begin{aligned} \|U_2U_2^TF_k\| &= \|F_k + \tilde{J}_k\tilde{d}_k\| \\ &\leq \|F_k + \tilde{J}_k(\bar{x}_k - x_k)\| \\ &\leq \|F_k + J_k(\bar{x}_k - x_k)\| + \|(\tilde{J}_k - J_k)(\bar{x}_k - x_k)\| \\ &\leq L\|\bar{x}_k - x_k\|^2 + \|(U_2\Sigma_2V_2^T)(\bar{x}_k - x_k)\| \\ &\leq L\|\bar{x}_k - x_k\|^2 + L\|x_k - x^*\|\|\bar{x}_k - x_k\| \\ &= o(\|\bar{x}_k - x_k\|). \end{aligned} \tag{3.13}$$

From the SVD of J_k, it is easy to get

$$d_k = -V_1(\Sigma_1^2 + \lambda_kI)^{-1}\Sigma_1U_1^TF_k - V_2(\Sigma_2^2 + \lambda_kI)^{-1}\Sigma_2U_2^TF_k.$$

Then we know

$$\begin{aligned} \|F_k + J_kd_k\| &= \|\lambda_kU_1(\Sigma_1^2 + \lambda_kI)^{-1}U_1^TF_k + \lambda_kU_2(\Sigma_2^2 + \lambda_kI)^{-1}U_2^TF_k\| \\ &\leq \lambda_k\|\Sigma_1^{-2}\|\|F_k\| + \|U_2U_2^TF_k\| \\ &\leq L^2\mu C_3^2\|\bar{x}_k - x_k\|^2 + \|U_2U_2^TF_k\| \\ &= o(\|\bar{x}_k - x_k\|), \end{aligned} \tag{3.14}$$

where the second inequality uses (3.5), (3.3), $F(\bar{x}_k) = 0$ and (3.13).

Note that $y_k = x_k + d_k$, then from (3.4) and (3.14), we have

$$\begin{aligned} \|F(y_k)\| &\leq \|F(y_k) - F_k - J_kd_k\| + \|F_k + J_kd_k\| \\ &\leq L\|d_k\|^2 + o(\|\bar{x}_k - x_k\|) \\ &= o(\|\bar{x}_k - x_k\|). \end{aligned}$$

□

Lemma 3.5. *Let Assumption 3.1 hold. Then for sufficiently large k, we have $\alpha_k \equiv 1$.*

Proof. From (1.5), it is easy to see that $\hat{d}_k$ is the minimizer of the convex optimization problem:

$$\min_{d\in R^n} \varphi_{k,2}(d) = \|F(y_k) + J_kd\|^2 + \lambda_k\|d\|^2.$$

Then by Lemma 3.4, we have

$$\|F(y_k) + J_k\hat{d}_k\| \leq \sqrt{\varphi_{k,2}(\hat{d}_k)} \leq \sqrt{\varphi_{k,2}(0)} = \|F(y_k)\| = o(\|\bar{x}_k - x_k\|). \tag{3.15}$$

Therefore from (3.4), Lemma 3.3, Lemma 3.4, (3.14) and (3.15), we have

$$\begin{aligned} &\|F(x_k + d_k + \hat{d}_k)\| \\ \leq\ &\|F(x_k + d_k + \hat{d}_k) - F_k - J_k(d_k + \hat{d}_k)\| + \|F_k + J_k(d_k + \hat{d}_k)\| \\ \leq\ &C_6(\|d_k\|^2 + \|\hat{d}_k\|^2) + \|F_k + J_kd_k\| + \|F(y_k) + J_k\hat{d}_k\| + \|F(y_k)\| \\ =\ &o(\|\bar{x}_k - x_k\|) \\ =\ &\eta_k\|\bar{x}_k - x_k\| \leq c_1^{-1}\eta_k\|F_k\| \leq \rho\|F_k\|, \end{aligned}$$

for some constant $C_6 > 0$, where the last line uses the error bound condition and $\eta_k \to 0$.

The above inequalities show that (1.6) holds for all sufficiently large k, which means $\alpha_k \equiv 1$ for sufficiently large k, i.e., the unit step will be always accepted finally. □

Then cubic convergence of Algorithm 1.1 can be established using completely same arguments as [4]. We list this local convergence result but omit the proof here.

Theorem 3.1. *Let Assumption 3.1 hold. Then $x_k \to x^*$ cubically.*

Acknowledgments. We would like to thank the referees whose very helpful suggestions led to much improvement of this paper.

References

[1] W. Cruz, J. M. Martinez and M. Raydan, *Special residual method without gradient information for solving large scale nonlinear systems of equations*, Math. Comput., 75 (2006), pp. 1429-1448.

[2] J. E. Dennis and J. J. More, *A characterization of superlinear convergence and its applications to quasi-Newton methods*, Math. Comput., 28 (1974), pp. 549-560.

[3] J. E. Dennis, R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

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

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

[6] D. Goldfarb, *Curvilinear path steplength algorithms for minimization which use directions of negative curvature*, Math. Program., 18 (1980), pp. 31-40.

[7] L. Grippo, F. Lampariello and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM J. Numer. Anal., 23 (1986), pp. 707-716.

[8] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, 1995.

[9] C. T. Kelley, *Iterative Methods for Optimization (Frontiers in Applied Mathematics 18)*, SIAM, Phaladelphia, 1999.

[10] D. Li and M. Fukushima, *A globally and superlinearly convergent Gauss-Newton-based BFGS method for symmetric nonlinear equations*, SIAM J. Numer. Anal., 37 (1999), pp. 152-172.

[11] D. Li and M. Fukushima, *A derivative-free line search and global convergence of Broyden-like method for nonlinear equations*, Optim. Methods Softw., 13 (2000), pp. 181-201.

[12] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[13] G. W. Stewart and J. G. Sun, *Matrix Perturbation Theory*, Academic Press, San Diego, CA, 1990.

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

[15] N. Yamashita and M. Fukushima, *On the rate of convergence of the Levenberg-Marquardt method*, Computing (Supp.), 15 (2001), pp. 237-249.

[16] W. Zhou and X. Chen, *Global convergence of a new hybrid Gauss-Newton structured BFGS methods for nonlinear least squares problems*, SIAM J. Optim., 20 (2010), pp. 2422-2441.

[17] W. Zhou and D. Li, *A globally convergent BFGS method for nonlinear monotone equations without any merit functions*, Math. Comput., 77 (2008), pp. 2231-2240.