Journal of Scientific Computing (2019) 78:531–548
https://doi.org/10.1007/s10915-018-0777-8

CrossMark

An Adaptive Multi-step Levenberg–Marquardt Method

Jinyan Fan[1] · Jianchao Huang[2] · Jianyu Pan[3]

Received: 20 March 2017 / Revised: 22 June 2018 / Accepted: 23 June 2018 / Published online: 28 June 2018

Abstract
We propose an adaptive multi-step Levenberg–Marquardt (LM) method for nonlinear equations. The adaptive scheme can decide automatically whether an iteration should evaluate the Jacobian matrix at the current iterate to compute an LM step, or use the latest evaluated Jacobian to compute an approximate LM step, so that not only the Jacobian evaluation but also the linear algebra work can be saved. It is shown that the adaptive multi-step LM method converges superlinearly under the local error bound condition, which does not require the full column rank of the Jacobian at the solution. Numerical experiments demonstrate the efficiency of the adaptive multi-step LM method.

Keywords Nonlinear equations · Levenberg–Marquardt method · Trust region method

Mathematics Subject Classification 65K05 · 65K10 · 90C30

1 Introduction

We consider the system of nonlinear equations

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

where $F(x) : R^n \rightarrow R^m$ is a continuously differentiable function. Nonlinear equations have wide applications in technology, mechanics, economy and so on. For example, physical models that are expressed as nonlinear partial differential equations become systems of nonlinear equations when discretized [5].

The first author was supported in part by NSFC Grant 11571234. The third author was supported in part by NSFC Grant 11371145 and 11771148, and Science and Technology Commission of Shanghai Municipality Grant 13dz2260400..

✉ Jinyan Fan
jyfan@sjtu.edu.cn

1 School of Mathematical Sciences, and Key Lab of Scientific and Engineering Computing (Ministry of Education), Shanghai Jiao Tong University, Shanghai 200240, China

2 School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China

3 School of Mathematical Sciences, Shanghai Key Laboratory of PMMP, East China Normal University, Shanghai 200241, China

It is natural to transform the nonlinear equations (1.1) to the nonlinear least squares problem

$$\min_{x \in R^n} \|F(x)\|^2, \tag{1.2}$$

where $\|\cdot\|$ refers to the 2-norm. Obviously, (1.1) has a solution if and only if the minimal value of (1.2) is zero. The Gauss–Newton method is the most well-known method for (1.2). At the k-th iteration, it computes the Gauss–Newton step

$$d_k^{GN} = -(J_k^T J_k)^{-1} J_k^T F_k, \tag{1.3}$$

where $F_k = F(x_k)$ and $J_k = F'(x_k)$ is the Jacobian at x_k. However, when the Jacobian is not of full rank, the Gauss–Newton step is not well-defined. To overcome this difficulty, the Levenberg–Marquardt (LM) method introduces a positive parameter $\lambda_k > 0$ and computes the LM step

$$d_k = -(J_k^T J_k + \lambda_k I)^{-1} J_k^T F_k. \tag{1.4}$$

Both the Gauss–Newton method and the Levenberg–Marquardt method have quadratic convergence for (1.1) under the assumptions that the Jacobian $J(x)$ is Lipschitz continuous, $J(x)^T J(x)$ is nonsingular at the solution and the LM parameter is chosen suitably. Yamashita and Fukushima [13] proved that if the LM parameter is chosen as $\lambda_k = \|F_k\|^2$, then the LM method converges quadratically for (1.1) under the local error bound condition, which is weaker than the nonsingularity condition. Fan and Yuan [4] took $\lambda_k = \|F_k\|^\delta$ and showed that the LM method preserves the quadratic convergence for any $\delta \in [1, 2]$ under the local error bound condition.

The LM method evaluates the Jacobian at every iteration. When $F(x)$ is complicated or n is large, the cost of Jacobian evaluations may be expensive. To save the Jacobian evaluations as well as the linear algebra work, Fan [2] proposed a modified LM method. At every iteration, it uses the evaluated Jacobian to compute not only an LM step but also an approximate LM step. More generally, a multi-step LM method was given in [3]. At the k-th iteration, it computes one LM step and $p - 1$ approximate LM steps as follows:

$$d_{k,i} = -(J_k^T J_k + \lambda_k I)^{-1} J_k^T F(x_{k,i}), \quad i = 0, \ldots, p-1, \tag{1.5}$$

where $p \geq 1$ is a given integer and $x_{k,i} = x_{k,i-1} + d_{k,i-1}$ with $x_{k,0} = x_k$, then set the trial step as

$$d_k = \sum_{i=0}^{p-1} d_{k,i}. \tag{1.6}$$

That is, the multi-step LM method uses J_k as an approximation of $J(x_{k,i})$ to compute the approximate LM step $d_{k,i}$. Hence, the Jacobian evaluation and the matrix factorization are done after every p computations of the step. Counting as a complete iteration of the full p steps between the Jacobian evaluation and the matrix factorization, the multi-step LM method converges with Q-order $p + 1$ under the local error bound condition.

It looks like that the bigger the p, the more Jacobian evaluations could be saved. However, the bigger p, meanwhile, implies the possibly worse approximation of the Jacobian at the iterate. On the other hand, different problems may prefer to different p. But, the best p for a problem is usually not known. These observations motivate us to develop an adaptive scheme to automatically decide whether an iteration should evaluate the Jacobian at the current iterate to compute an LM step or use the latest evaluated Jacobian to compute an approximate LM step. That is, the number of approximate LM steps could vary at different iterations, but it is at most $t - 1$ to avoid worse approximation of the Jacobian, where $t \geq 1$ is a given integer.

Counting every LM step and every approximate LM step as a single iteration, in this paper, we compute the trial step d_k by solving

$$\left(G_k^T G_k + \lambda_k I\right) d = - G_k^T F_k, \tag{1.7}$$

where G_k is the Jacobian J_k or the Jacobian used at the last iteration. If d_k is satisfactory and less than $t - 1$ approximate LM steps are computed, we regard G_k as a good approximation of the Jacobian at the current iterate, and use it to compute another approximate LM step at next iteration.

Define the ratio of the actual reduction to the predicted reduction of the merit function $\|F(x)\|^2$ at the k-th iteration as

$$r_k = \frac{Ared_k}{Pred_k} = \frac{\|F_k\|^2 - \|F(x_k + d_k)\|^2}{\|F_k\|^2 - \|F_k + G_k d_k\|^2}. \tag{1.8}$$

The ratio r_k exploits the most relevant information on the step's quality at the iterate. It plays an important role in deciding whether d_k is acceptable. If r_k is positive (i.e., the iteration is successful), we accept d_k. Usually, we set

$$x_{k+1} = \begin{cases} x_k + d_k, & \text{if} \quad r_k \geq p_0, \\ x_k, & \text{otherwise}, \end{cases} \tag{1.9}$$

where p_0 is a small positive constant. Suppose $s - 1$ approximate LM steps that used the latest evaluated Jacobian have been computed. If the iteration is very successful and s is less than t, we regard the latest evaluated Jacobian as a good approximation at the iterate and keep it, otherwise, we evaluate the exact one. That is, we set

$$G_{k+1} = \begin{cases} G_k, & \text{if} \quad r_k \geq p_1 \text{ and } s < t, \\ J_{k+1}, & \text{otherwise}, \end{cases} \tag{1.10}$$

where $0 < p_0 < p_1 < 1$. Meanwhile, we update the LM parameter as follows:

$$\lambda_{k+1} = \begin{cases} \lambda_k, & \text{if} \quad r_k \geq p_1 \text{ and } s < t, \\ \mu_{k+1} \|F_{k+1}\|^\delta, & \text{otherwise}, \end{cases} \tag{1.11}$$

where

$$\mu_{k+1} = \begin{cases} c_1 \mu_k, & \text{if} \quad r_k < p_2, \\ \mu_k, & \text{if} \quad p_2 \leq r_k \leq p_3, \\ \max\{c_2 \mu_k, \mu_{\min}\}, & \text{if} \quad r_k > p_3. \end{cases} \tag{1.12}$$

Here $0 < c_2 < 1 < c_1$, $0 < p_0 < p_2 < p_1 < p_3 < 1$, $1 \leq \delta \leq 2$ and $\mu_{\min} > 0$ are positive constants. Based on (1.7)–(1.12), we propose an adaptive multi-step LM method for nonlinear equations (1.1) and show it converges superlinearly under the local error bound condition.

The paper is organized as follows. In Sect. 2, we propose an adaptive multi-step LM algorithm for (1.1). It is shown that the algorithm converges globally under certain assumptions. In Sect. 3, we discuss the convergence rate of the algorithm under the local error bound condition. Some numerical results are given in Sect. 4. Finally, we conclude the paper in Sect. 5.

2 An Adaptive Multi-step LM Algorithm and Global Convergence

In this section, we first propose the adaptive multi-step LM algorithm, then show it converges globally under certain assumptions.

The adaptive multi-step LM algorithm is presented as follows.

Algorithm 1: An adaptive multi-step Levenberg–Marquardt algorithm

Input: $x_1 \in R^n, c_1 > 1 > c_2 > 0, 0 < p_0 < p_2 < p_1 < p_3 < 1, 1 \le \delta \le 2, t \ge 1, \mu_1 > \mu_{\min} > 0$.

Set $G_1 = J_1, \lambda_1 = \mu_1 \|F_1\|^\delta, k := 1, s := 1, i := 1, k_i = 1$.

while $\|G_{k_i}^T F_{k_i}\| \neq 0$ **do**

  Compute d_k by solving (1.7).

  Compute $r_k = Ared_k / Pred_k$ by (1.8), and set x_{k+1} by (1.9).

  Update G_{k+1}, λ_{k+1} and μ_{k+1} by (1.10), (1.11) and (1.12), respectively.

  Set $k := k + 1$. If G_k is the Jacobian at x_k, set $s := 1, i := i + 1, k_i = k$, otherwise set $s := s + 1$.

end

We denote by $\bar{S} = \{k_i : i = 1, 2, \ldots\}$ the set of numbers at which iterations the Jacobians $J(x_{k_i})(i = 1, 2, \ldots)$ are used to compute LM steps. Let

$$s_i = k_{i+1} - k_i. \tag{2.1}$$

Since at most $t - 1$ approximate LM steps are computed, we have $s_i \le t$.

For any k, there exist k_i and $0 \le q \le s_i - 1$ such that

$$k = k_i + q. \tag{2.2}$$

Note that

$$G_{k_i} = G_{k_i+1} = \cdots = G_{k_i+s_i-1} = J_{k_i}, \tag{2.3}$$

thus the linear equations (1.7) can also be written as

$$\left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right) d = -\, J_{k_i}^T F_k \tag{2.4}$$

for $k = k_i, \ldots, k_i + s_i - 1$.

It can be easily checked that d_k is not only the minimizer of the convex minimization problem

$$\min_{d \in R^n} \|F_k + G_k d\|^2 + \lambda_k \|d\|^2 \triangleq \varphi_k(d), \tag{2.5}$$

but also the solution of the trust region problem

$$\begin{aligned} &\min_{d \in R^n} \ \|F_k + G_k d\|^2 \\ &\ s.t. \ \ \|d\| \le \Delta_k \triangleq \|d_k\|. \end{aligned} \tag{2.6}$$

Due to Powell's result [10, Theorem 4], we have the following lemma.

Lemma 2.1 *Let d_k be computed by* (1.7)*, then*

$$\|F_k\|^2 - \|F_k + G_k d_k\|^2 \ge \|G_k^T F_k\| \min\left\{\|d_k\|, \frac{\|G_k^T F_k\|}{\|G_k^T G_k\|}\right\}. \tag{2.7}$$

Lemma 2.1 indicates that the predicted reduction of the merit function is always nonnegative. It plays a crucial role in guaranteeing the global convergence of Algorithm 1. In the following, we show that, from the optimization point, at least one of the accumulation points of the sequence generated by Algorithm 1 is a stationary point of the merit function $\|F(x)\|^2$.

Theorem 2.2 *Suppose $F(x)$ is continuously differentiable, both $F(x)$ and $J(x)$ are Lipschitz continuous, i.e., there exist positive constants κ_{lf} and κ_{lj} such that*

$$\|F(y) - F(x)\| \le \kappa_{lf}\|y - x\|, \quad \forall x, y \in R^n, \tag{2.8}$$

and

$$\|J(y) - J(x)\| \le \kappa_{lj}\|y - x\|, \quad \forall x, y \in R^n. \tag{2.9}$$

Then, we have

$$\liminf_{k\to\infty} \|J_k^T F_k\| = 0. \tag{2.10}$$

Proof Suppose that (2.10) is not true. Then, there exists a constant $\varepsilon > 0$ such that

$$\left\|J_k^T F_k\right\| \ge \varepsilon, \quad \forall k. \tag{2.11}$$

By (2.8) and (2.9), we have

$$\|J(x)\| \le \kappa_{lf}, \quad \forall x \in R^n \tag{2.12}$$

and

$$\|F(y) - F(x) - J(x)(y - x)\| \le \kappa_{lj}\|y - x\|^2, \quad \forall x, y \in R^n. \tag{2.13}$$

Denote the index set of successful iterations by

$$S = \{k : r_k \ge p_0\}. \tag{2.14}$$

We consider S in two cases.

Case 1: S is finite. Then, there exists $\tilde{k}$ such that $r_k < p_0 < p_1$ for all $k \ge \tilde{k}$. Hence, $G_k = J_k$ for all $k \ge \tilde{k}$. Since $\{\|F_k\|\}$ is nonincreasing, by (1.12), $\mu_{k+1} = c_1\mu_k$ with $c_1 > 1$. So, $\mu_k \to +\infty$.

By (2.11) and (2.12), for $k \ge \tilde{k}$,

$$\|G_k^T F_k\| \le \|G_k\|\|F_k\| \le \kappa_{lf}\|F_1\| \tag{2.15}$$

and

$$\|F_k\| \ge \frac{\|G_k^T F_k\|}{\|G_k\|} \ge \frac{\|J_k^T F_k\|}{\kappa_{lf}} \ge \frac{\varepsilon}{\kappa_{lf}}. \tag{2.16}$$

This, together with $\lambda_k = \mu_k\|F_k\|^\delta$ and $\mu_k \to +\infty$, gives $\lambda_k \to +\infty$. Hence, by the definition of d_k, we get $d_k \to 0$.

Case 2: S is infinite. It follows from (2.12) and Lemma 2.1 that

$$\begin{aligned} \|F_1\|^2 &\geq \sum_{k\in S}\left(\|F_k\|^2-\|F_{k+1}\|^2\right) \\ &\geq \sum_{k\in S} p_0 Pred_k \geq \sum_{k\in S\cap\bar{S}} p_0 Pred_k \\ &\geq \sum_{k\in S\cap\bar{S}} p_0\|J_k^T F_k\| \min\left\{\|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\} \\ &\geq \sum_{k\in S\cap\bar{S}} p_0\varepsilon \min\left\{\|d_k\|, \frac{\varepsilon}{\kappa_{lf}^2}\right\}. \end{aligned} \tag{2.17}$$

We claim that the set $S\cap\bar{S}$ is infinite. Suppose to the contrary that it is finite. Let $k_{\bar{i}}$ be the largest index in it. Then, for $i>\bar{i}$, $r_{k_i}<p_0<p_1$. By (1.10), $G_{k_i+1}=J_{k_i+1}$. Thus, $k_{i+1}=k_i+1$. Moreover, $k_{i+1}\notin S$. Deducing by induction, we obtain that $r_k<p_0$ for all sufficiently large k, which contradicts to the infinity of S. Hence, $S\cap\bar{S}$ is infinite.

Note that $k_i\in\bar{S}$ according to the definition of $\bar{S}$. By (2.17), we obtain $d_{k_i}\to 0$ for $k_i\in S$. Since $d_{k_i}=0$ for $k_i\notin S$, we have $d_{k_i}\to 0$. This, together with $\|G_{k_i}\|\leq\kappa_{lf}$, $\|G_{k_i}^T F_{k_i}\|=\|J_{k_i}^T F_{k_i}\|\geq\varepsilon$ and (1.7), gives $\lambda_{k_i}\to+\infty$.

By (2.12) and (2.13), for $k=k_i+1,\ldots,k_i+s_i-1$,

$$\begin{aligned} \|d_k\| &= \left\|-(G_k^T G_k+\lambda_k I)^{-1}G_k^T F_k\right\| \\ &\leq \left\|(J_{k_i}^T J_{k_i}+\lambda_{k_i}I)^{-1}J_{k_i}^T F_{k_i}\right\| + \left\|(J_{k_i}^T J_{k_i}+\lambda_{k_i}I)^{-1}J_{k_i}^T J_{k_i}\left(\sum_{j=k_i}^{k-1} d_j\right)\right\| \\ &\quad + \kappa_{lj}\left\|(J_{k_i}^T J_{k_i}+\lambda_{k_i}I)^{-1}J_{k_i}^T\right\| \left\|\left(\sum_{j=k_i}^{k-1} d_j\right)\right\|^2 \\ &\leq \|d_{k_i}\| + \sum_{j=k_i}^{k-1}\|d_j\| + \frac{\kappa_{lf}\kappa_{lj}}{\lambda_{k_i}}\left(\sum_{j=k_i}^{k-1}\|d_j\|\right)^2. \end{aligned} \tag{2.18}$$

Since $\lambda_{k_i}\to+\infty$, we have

$$\|d_{k_i+1}\|\leq 3\|d_{k_i}\| \tag{2.19}$$

and

$$\begin{aligned} \|d_{k_i+2}\| &\leq \|d_{k_i}\|+\|d_{k_i}\|+\|d_{k_i+1}\|+\frac{\kappa_{lf}\kappa_{lj}}{\lambda_{k_i}}\left(\|d_{k_i}\|+\|d_{k_i+1}\|\right)^2 \\ &\leq 21\|d_{k_i}\| \end{aligned} \tag{2.20}$$

for sufficiently large k_i. By induction, there exists a positive constant $\hat{c}$ such that, for $k=k_i+1,\ldots,k_i+s_i-1$,

$$\|d_k\|\leq\hat{c}\|d_{k_i}\| \tag{2.21}$$

holds for all sufficiently large k_i. Since $s_i \le t$ for all i, we have $d_k \to 0$ and

$$\|J_k - G_k\| = \|J_k - J_{k_i}\| \le \kappa_{lj} \sum_{j=k_i}^{k-1} \|d_j\| \to 0. \tag{2.22}$$

Hence, by (2.11),

$$\left\|G_k^T F_k\right\| \ge \left\|J_k^T F_k\right\| - \left\|(J_k - G_k)^T F_k\right\| \ge \left\|J_k^T F_k\right\| - \|J_k - G_k\|\|F_1\| \ge \varepsilon/2 \tag{2.23}$$

holds for sufficiently large k. Since $\|F_k\| \ge \left\|G_k^T F_k\right\| / \|G_k\| \ge \frac{\varepsilon}{2\kappa_{lf}}$, by the definition of d_k and λ_k, we have $\lambda_k \to +\infty$ and $\mu_k \to +\infty$.

Therefore, no matter S is infinite or not, we obtain

$$d_k \to 0, \quad \lambda_k \to +\infty, \quad \mu_k \to +\infty, \tag{2.24}$$

moreover, (2.22) and (2.23) hold true. It then follows from (2.12), (2.13), (2.22), (2.23), Lemma 2.1 and $\|F_k\| \le \|F_1\|$ that

$$\begin{aligned}
|r_k - 1| &= \left|\frac{Ared_k - Pred_k}{Pred_k}\right| = \left|\frac{\|F_{k+1}\|^2 - \|F_k + G_k d_k\|^2}{\|F_k\|^2 - \|F_k + G_k d_k\|^2}\right| \\
&\le \frac{\left|\|F_k + J_k d_k\|^2 - \|F_k + G_k d_k\|^2\right| + 2\kappa_{lj}\,\|F_k + J_k d_k\|\,\|d_k\|^2 + \kappa_{lj}^2\,\|d_k\|^4}{\left\|G_k^T F_k\right\| \min\left\{\|d_k\|, \frac{\left\|G_k^T F_k\right\|}{\left\|G_k^T G_k\right\|}\right\}} \\
&\le \frac{\left(\left\|J_k^T J_k\right\| + \left\|G_k^T G_k\right\|\right)\|d_k\|^2 + 2\,\|J_k - G_k\|\,\|F_k\|\,\|d_k\|}{\frac{\varepsilon}{2} \min\left\{\|d_k\|, \frac{\varepsilon}{2\kappa_{lf}^2}\right\}} \\
&\quad + \frac{2\kappa_{lj}\|F_1\|\|d_k\|^2 + 2\kappa_{lj}\kappa_{lf}\|d_k\|^3 + \kappa_{lj}^4\|d_k\|^4}{\frac{\varepsilon}{2} \min\left\{\|d_k\|, \frac{\varepsilon}{2\kappa_{lf}^2}\right\}} \\
&\to 0.
\end{aligned} \tag{2.25}$$

Thus, $r_k \to 1$. So, by (1.10)–(1.12), there exists a positive $\hat{\mu}$ such that $\mu_k \le \hat{\mu}$ holds for all sufficiently large k, which is a contradiction to (2.24). Hence, (2.11) can not be true. Therefore, we obtain (2.10). □

3 Local Convergence

In this section, we first investigate the properties of the trial step, then show that the LM parameter is bounded, finally we prove that Algorithm 1 converges superlinearly under the local error bound condition. We make the following assumptions.

Assumption 3.1 The sequence $\{x_k\}$ generated by Algorithm 1 satisfies $\text{dist}(x_k, X^*) \to 0$, and there exist $x^* \in X^*$ and $0 < r < 1$ such that $\|x_k - x^*\| \le r/2$ for all large k.

Assumption 3.2 $F(x)$ is continuously differentiable, and $J(x)$ is Lipschitz continuous on $N(x^*, r) = \{x : \|x - x^*\| \le r\}$, i.e., there exists a constant $\kappa_{lj} > 0$ such that

$$\|J(y) - J(x)\| \le \kappa_{lj}\|y - x\|, \quad \forall x, y \in N(x^*, r). \tag{3.1}$$

Assumption 3.3 $\|F(x)\|$ provides a local error bound on $N(x^*, r)$, i.e., there exists a constant $\kappa_{leb} > 0$ such that

$$\|F(x)\| \geq \kappa_{leb} \operatorname{dist}(x, X^*), \quad \forall x \in N(x^*, r). \tag{3.2}$$

By Assumption 3.2,

$$\|F(y) - F(x) - J(x)(y-x)\| \leq \kappa_{lj}\|y-x\|^2, \quad \forall x, y \in N(x^*, r), \tag{3.3}$$

and there exists a constant $\kappa_{lf} > 0$ such that

$$\|F(y) - F(x)\| \leq \kappa_{lf}\|y-x\|, \quad \forall x, y \in N(x^*, r). \tag{3.4}$$

Denote by $\bar{x}_k$ the vector in X^* that satisfies

$$\|\bar{x}_k - x_k\| = \operatorname{dist}(x_k, X^*). \tag{3.5}$$

Then, for all large k,

$$\|\bar{x}_k - x_k\| \leq \|x^* - x_k\| \leq \frac{r}{2} \tag{3.6}$$

and

$$\|\bar{x}_k - x^*\| \leq \|\bar{x}_k - x_k\| + \|x^* - x_k\| \leq r, \tag{3.7}$$

so $\bar{x}_k \in N(x^*, r)$. In the following, the iterations k we consider are large.

3.1 Properties of the Step d_k

In this subsection, we show the relationship between the size of d_k and the distance from x_{k_i} to the solution set.

Lemma 3.4 *Under Assumptions 3.1–3.3, there exists a constant $c > 0$ such that*

$$\|d_k\| \leq c\|\bar{x}_{k_i} - x_{k_i}\|, \quad k = k_i, \ldots, k_i + s_i - 1. \tag{3.8}$$

Proof It follows from (1.12) and (3.2) that

$$\lambda_{k_i} = \mu_{k_i}\|F_{k_i}\|^\delta \geq \mu_{\min}\kappa_{leb}^\delta\|\bar{x}_{k_i} - x_{k_i}\|^\delta. \tag{3.9}$$

Since d_{k_i} is the minimizer of $\varphi_{k_i}(d)$, by (3.6), (3.3), $0 < r < 1$ and $1 \leq \delta \leq 2$, we have

$$\begin{aligned}
\|d_{k_i}\|^2 &\leq \frac{\varphi_{k_i}(d_{k_i})}{\lambda_{k_i}} \\
&\leq \frac{\varphi_{k_i}(\bar{x}_{k_i} - x_{k_i})}{\lambda_{k_i}} \\
&= \frac{\|F_{k_i} + J_{k_i}(\bar{x}_{k_i} - x_{k_i})\|^2}{\lambda_{k_i}} + \|\bar{x}_{k_i} - x_{k_i}\|^2 \\
&\leq \frac{\kappa_{lj}^2}{\mu_{\min}\kappa_{leb}^\delta}\|\bar{x}_{k_i} - x_{k_i}\|^{4-\delta} + \|\bar{x}_{k_i} - x_{k_i}\|^2 \\
&\leq c_0\|\bar{x}_{k_i} - x_{k_i}\|,
\end{aligned} \tag{3.10}$$

where $c_0 = \sqrt{\mu_{\min}^{-1}\kappa_{leb}^{-\delta}\kappa_{lj}^2 + 1}$.

For $k = k_i + 1, \ldots, k_i + s_i - 1$, by (3.3),

$$\begin{aligned}\|d_k\| &= \left\| - \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right)^{-1} J_{k_i}^T F_k \right\| \\ &\leq \left\| - \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right)^{-1} J_{k_i}^T F_{k_i} \right\| + \left\| - \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right)^{-1} J_{k_i}^T J_{k_i} \left(\sum_{j=k_i}^{k-1} d_j\right) \right\| \\ &\quad + \kappa_{lj} \left\| - \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right)^{-1} J_{k_i}^T \right\| \left\| \sum_{j=k_i}^{k-1} d_j \right\|^2 \\ &\leq \|d_{k_i}\| + \sum_{j=k_i}^{k-1} \|d_j\| + \kappa_{lj} \left\| \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right)^{-1} J_{k_i}^T \right\| \left(\sum_{j=k_i}^{k-1} \|d_j\|\right)^2 . \end{aligned} \tag{3.11}$$

By (3.9),

$$\begin{aligned}\left\| \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right)^{-1} J_{k_i}^T \right\| &= \left\| \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right)^{-1} J_{k_i}^T J_{k_i} \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right)^{-1} \right\|^{1/2} \\ &\leq \left\| \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right)^{-1} \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right) \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right)^{-1} \right\|^{1/2} \\ &= \left\| \left(J_{k_i}^T J_{k_i} + \lambda_{k_i} I\right)^{-1} \right\|^{1/2} \\ &\leq \frac{1}{\sqrt{\lambda_{k_i}}} \leq \frac{1}{\mu_{\min}^{\frac{1}{2}} \kappa_{leb}^{\frac{\delta}{2}}} \|\bar{x}_{k_i} - x_{k_i}\|^{-\frac{\delta}{2}} . \end{aligned} \tag{3.12}$$

So, it follows from (3.6), (3.10) and $1 \leq \delta \leq 2$ that

$$\begin{aligned}\|d_{k_i+1}\| &\leq \|d_{k_i}\| + \|d_{k_i}\| + \kappa_{lj} \mu_{\min}^{-\frac{1}{2}} \kappa_{leb}^{-\frac{\delta}{2}} c_0 \|d_{k_i}\| \|\bar{x}_{k_i} - x_{k_i}\|^{1-\frac{\delta}{2}} \\ &\leq 2\|d_{k_i}\| + \bar{c}\|d_{k_i}\| \\ &= c_1 \|d_{k_i}\|, \end{aligned} \tag{3.13}$$

where $\bar{c} = \kappa_{lj} \mu_{\min}^{-\frac{1}{2}} \kappa_{leb}^{-\frac{\delta}{2}} c_0$ and $c_1 = 2 + \bar{c}$ are positive constants. Similarly,

$$\begin{aligned}\|d_{k_i+2}\| &\leq \|d_{k_i}\| + \|d_{k_i}\| + \|d_{k_i+1}\| + \kappa_{lj} \mu_{\min}^{-\frac{1}{2}} \kappa_{leb}^{-\frac{\delta}{2}} c_0 (1 + c_1)^2 \|d_{k_i}\| \|\bar{x}_{k_i} - x_{k_i}\|^{1-\frac{\delta}{2}} \\ &\leq \left(2 + c_1 + \bar{c}(1 + c_1)^2\right) \|d_{k_i}\| \\ &= c_2 \|d_{k_i}\|, \end{aligned} \tag{3.14}$$

where $c_2 = 2 + c_1 + \bar{c}(1 + c_1)^2$. Let

$$c_j = 2 + c_1 + \cdots + c_{j-1} + \bar{c}(1 + c_1 + \cdots + c_{j-1})^2, \quad j = 3, \ldots, t-1. \tag{3.15}$$

Then, by induction, we have

$$\|d_{k_i+q}\| \leq c_q \|d_{k_i}\|, \quad q = 3, \ldots, s_i - 1. \tag{3.16}$$

Since $s_i \leq t$ for all i, $c_{t-1} > c_{t-2} > \cdots > c_2 > c_1 > 1$, by (3.10), we obtain (3.8), where $c = c_0 c_{t-1}$. □

3.2 Boundedness of the LM Parameter

The updating rule of μ_k indicates that μ_k is bounded below. In the following, we show that μ_k is bounded above.

Lemma 3.5 *Under Assumptions 3.1–3.3, there exists a constant $\mu_{\max} > 0$ such that*

$$\mu_k \leq \mu_{\max} \tag{3.17}$$

holds for all sufficiently large k.

Proof We first prove that

$$\|F_k\|^2 - \|F_k + G_k d_k\|^2 \geq \frac{\kappa_{leb}}{2} \|F_k\| \min\{\|d_k\|, \|\bar{x}_k - x_k\|\} \tag{3.18}$$

holds for sufficiently large k.

It follows from (3.1), (3.3), Lemma 3.4 and $s_i \leq t$ that, for $k = k_i, k_i + 1, \ldots, k_i + s_i - 1$,

$$\begin{aligned}\|F_k + G_k(\bar{x}_k - x_k)\| &\leq \|F_k + J_k(\bar{x}_k - x_k)\| + \|G_k - J_k\| \|\bar{x}_k - x_k\| \\ &\leq \kappa_{lj} \|\bar{x}_k - x_k\|^2 + \kappa_{lj} \left(\sum_{j=k_i}^{k-1} \|d_j\| \right) \|\bar{x}_k - x_k\| \\ &\leq \kappa_{lj} \|\bar{x}_k - x_k\|^2 + \kappa_{lj} ct \|\bar{x}_{k_i} - x_{k_i}\| \|\bar{x}_k - x_k\|. \end{aligned} \tag{3.19}$$

Note that $\|\bar{x}_k - x_k\| \to 0$ and $\|\bar{x}_{k_i} - x_{k_i}\| \to 0$, we obtain that

$$\|F_k + G_k(\bar{x}_k - x_k)\| \leq \frac{\kappa_{leb}}{2} \|\bar{x}_k - x_k\| \tag{3.20}$$

holds for sufficiently large k.

We consider in two cases.

Case 1: $\|\bar{x}_k - x_k\| \leq \|d_k\|$. Since d_k is a solution of the trust region problem (2.6), by (3.2) and (3.20),

$$\begin{aligned}\|F_k\| - \|F_k + G_k d_k\| &\geq \|F_k\| - \|F_k + G_k(\bar{x}_k - x_k)\| \\ &\geq \kappa_{leb} \|\bar{x}_k - x_k\| - \frac{\kappa_{leb}}{2} \|\bar{x}_k - x_k\| \geq \frac{\kappa_{leb}}{2} \|\bar{x}_k - x_k\| \end{aligned} \tag{3.21}$$

holds for sufficiently large k.

Case 2: $\|\bar{x}_k - x_k\| > \|d_k\|$. Similarly, by (3.2) and (3.20),

$$\begin{aligned}\|F_k\| - \|F_k + G_k d_k\| &\geq \|F_k\| - \left\| F_k + \frac{\|d_k\|}{\|\bar{x}_k - x_k\|} G_k(\bar{x}_k - x_k) \right\| \\ &\geq \frac{\|d_k\|}{\|\bar{x}_k - x_k\|} (\|F_k\| - \|F_k + G_k(\bar{x}_k - x_k)\|) \\ &\geq \frac{\|d_k\|}{\|\bar{x}_k - x_k\|} \frac{\kappa_{leb}}{2} \|\bar{x}_k - x_k\| \\ &\geq \frac{\kappa_{leb}}{2} \|d_k\|. \end{aligned} \tag{3.22}$$

Combining (3.21) with (3.22), we get

$$\begin{aligned}\|F_k\|^2 - \|F_k + G_k d_k\|^2 &\geq \frac{\kappa_{leb}}{2} (\|F_k\| + \|F_k + G_k d_k\|) \min\{\|d_k\|, \|\bar{x}_k - x_k\|\} \\ &\geq \frac{\kappa_{leb}}{2} \|F_k\| \min\{\|d_k\|, \|\bar{x}_k - x_k\|\}, \end{aligned} \tag{3.23}$$

which yields (3.18).

It then follows from (3.2), (3.3), Lemma 3.4 and $\|F_{k_i} + J_{k_i} d_{k_i}\| \le \|F_{k_i}\|$ that

$$|r_{k_i} - 1| = \left|\frac{Ared_{k_i} - Pred_{k_i}}{Pred_{k_i}}\right| = \left|\frac{\|F_{k_i+1}\|^2 - \|F_{k_i} + J_{k_i} d_{k_i}\|^2}{\|F_{k_i}\|^2 - \|F_{k_i} + J_{k_i} d_{k_i}\|^2}\right|$$
$$\le \left|\frac{2\|F_{k_i} + J_{k_i} d_{k_i}\|\|d_{k_i}\|^2 + \kappa_{lj}^2\|d_{k_i}\|^4}{\frac{\kappa_{leb}}{2}\|F_{k_i}\|\min\{\|d_{k_i}\|, \|\bar{x}_{k_i} - x_{k_i}\|\}}\right|$$
$$\to 0. \tag{3.24}$$

This implies that $r_{k_i} \to 1$. Note that for $k \notin \bar{S}$, $G_{k+1} = G_k$ and $r_k \ge p_1 > p_2$, so $\mu_{k+1} \le \mu_k$. By (1.10)–(1.12), there exists a constant $\mu_{\max} > 0$ such that (3.17) holds true. □

3.3 Convergence Rate of Algorithm 1

Due to the results given by Behling and Iusem in [1], we assume that $\operatorname{rank}(J(\bar{x})) = r$ for all $\bar{x} \in N(x^*, b) \cap X^*$. Suppose the SVD of $J(\bar{x}_{k_i})$ is

$$J(\bar{x}_{k_i}) = \bar{U}_{k_i}\bar{\Sigma}_{k_i}\bar{V}_{k_i}^T$$
$$= (\bar{U}_{k_i,1}, \bar{U}_{k_i,2})\begin{pmatrix}\bar{\Sigma}_{k_i,1} & \\ & 0\end{pmatrix}\begin{pmatrix}\bar{V}_{k_i,1}^T \\ \bar{V}_{k_i,2}^T\end{pmatrix}$$
$$= \bar{U}_{k_i,1}\bar{\Sigma}_{k_i,1}\bar{V}_{k_i,1}^T, \tag{3.25}$$

where $\bar{\Sigma}_{k_i,1} = \operatorname{diag}(\bar{\sigma}_{k_i,1}, \bar{\sigma}_{k_i,2}, \ldots, \bar{\sigma}_{k_i,r})$ with $\bar{\sigma}_{k_i,1} \ge \cdots \ge \bar{\sigma}_{k_i,r} > 0$. Suppose the SVD of $J(x_{k_i})$ is

$$J_{k_i} = U_{k_i}\Sigma_{k_i}V_{k_i}^T$$
$$= (U_{k_i,1}, U_{k_i,2})\begin{pmatrix}\Sigma_{k_i,1} & \\ & \Sigma_{k_i,2}\end{pmatrix}\begin{pmatrix}V_{k_i,1}^T \\ V_{k_i,2}^T\end{pmatrix}$$
$$= U_{k_i,1}\Sigma_{k_i,1}V_{k_i,1}^T + U_{k_i,2}\Sigma_{k_i,2}V_{k_i,2}^T, \tag{3.26}$$

where $\Sigma_{k_i,1} = \operatorname{diag}(\sigma_{k_i,1}, \sigma_{k_i,2}, \ldots, \sigma_{k_i,r})$ with $\sigma_{k_i,1} \ge \cdots \ge \sigma_{k_i,r} > 0$ and $\Sigma_{k_i,2} = \operatorname{diag}(\sigma_{k_i,r+1}, \sigma_{k_i,r+2}, \ldots, \sigma_{k_i,n})$ with $\sigma_{k_i,r+1} \ge \cdots \ge \sigma_{k_i,n} \ge 0$.

By (3.1) and the theory of matrix perturbation [12], we obtain

$$\left\|\operatorname{diag}\left(\Sigma_{k_i,1} - \bar{\Sigma}_{k_i,1}, \Sigma_{k_i,2}\right)\right\| \le \|J_{k_i} - J(\bar{x}_{k_i})\| \le \kappa_{lj}\|\bar{x}_{k_i} - x_{k_i}\|. \tag{3.27}$$

Lemma 3.6 *Under Assumptions 3.1–3.3, there exist positive constants l_1 and l_2 such that*

$$\|d_k\| \le l_1\|\bar{x}_{k_i} - x_{k_i}\|^{k-k_i+1}, \quad k = k_i, \ldots, k_i + s_i - 1, \tag{3.28}$$
$$\|F_k + G_k d_k\| \le l_2\|\bar{x}_{k_i} - x_{k_i}\|^{k-k_i+2}, \quad k = k_i, \ldots, k_i + s_i - 1. \tag{3.29}$$

Proof We prove by induction. It was shown in [2] that the results hold true for $k = k_i$ and $k = k_i + 1$.

Suppose the results hold true for $k - 1 (k_i + 2 < k < k_i + s_i - 1)$, that is, there exist constants $\bar{l}_1$ and $\bar{l}_2$ such that

$$\|d_{k-1}\| \le \bar{l}_1\|\bar{x}_{k_i} - x_{k_i}\|^{k-k_i}, \tag{3.30}$$
$$\|F_{k-1} + G_{k-1}d_{k-1}\| \le \bar{l}_2\|\bar{x}_{k_i} - x_{k_i}\|^{k-k_i+1}. \tag{3.31}$$

It then follows from (3.1), (3.3) and Lemma 3.4 that

$$\begin{aligned}\|F_k\| &= \|F(x_{k-1}+d_{k-1})\| \\ &\le \|F_{k-1}+J_{k-1}d_{k-1}\| + \kappa_{lj}\|d_{k-1}\|^2 \\ &\le \|F_{k-1}+G_{k-1}d_{k-1}\| + \|(J_{k-1}-G_{k-1})\, d_{k-1}\| + \kappa_{lj}\|d_{k-1}\|^2 \\ &= \|F_{k-1}+G_{k-1}d_{k-1}\| + \left\|J_{k-1}-J_{k_i}\right\| \|d_{k-1}\| + \kappa_{lj}\|d_{k-1}\|^2 \\ &\le \bar{l}_2\|\bar{x}_{k_i}-x_{k_i}\|^{k-k_i+1} + \kappa_{lj}\sum_{j=k_i}^{k-2}\|d_j\|\|d_{k-1}\| + \kappa_{lj}\bar{l}_1^2\|\bar{x}_{k_i}-x_{k_i}\|^{2(k-k_i)} \\ &\le \bar{l}_3\|\bar{x}_{k_i}-x_{k_i}\|^{k-k_i+1}, \end{aligned} \tag{3.32}$$

where $\bar{l}_3 = \bar{l}_2 + \kappa_{lj}ct\bar{l}_1 + \kappa_{lj}\bar{l}_1^2$. Hence,

$$\left\|U_{k_i,1}U_{k_i,1}^T F_k\right\| \le \|F_k\| \le \bar{l}_3\|\bar{x}_{k_i}-x_{k_i}\|^{k-k_i+1}. \tag{3.33}$$

Moreover, by (3.2),

$$\|\bar{x}_k - x_k\| \le \kappa_{leb}^{-1}\|F_k\| \le \kappa_{leb}^{-1}\bar{l}_3\|\bar{x}_{k_i}-x_{k_i}\|^{k-k_i+1}. \tag{3.34}$$

Let $\tilde{J}_{k_i} = U_{k_i,1}\Sigma_{k_i,1}V_{k_i,1}^T$ and $\tilde{u}_k = -\tilde{J}_{k_i}^{+}F_k$. Then, $\tilde{u}_k$ is the least squares solution of the problem $\min\limits_{u\in R^n}\|F_k+\tilde{J}_{k_i}u\|$. By (3.1), (3.3) and (3.27),

$$\begin{aligned}\left\|U_{k_i,2}U_{k_i,2}^T F_k\right\| &= \|F_k + \tilde{J}_{k_i}\tilde{u}_k\| \le \|F_k + \tilde{J}_{k_i}(\bar{x}_k - x_k)\| \\ &\le \|F_k + J_k(\bar{x}_k-x_k)\| + \|J_{k_i}-J_k\|\|\bar{x}_k-x_k\| + \|\tilde{J}_{k_i}-J_{k_i}\|\|\bar{x}_k-x_k\| \\ &\le \kappa_{lj}\|\bar{x}_k-x_k\|^2 + \kappa_{lj}\sum_{j=k_i}^{k-1}\|d_j\|\|\bar{x}_k-x_k\| + \left\|U_{k_i,2}\Sigma_{k_i,2}V_{k_i,2}^T\right\| \|\bar{x}_k-x_k\| \\ &\le \bar{l}_4\|\bar{x}_{k_i}-x_{k_i}\|^{k-k_i+2}, \end{aligned} \tag{3.35}$$

where $\bar{l}_4 = \kappa_{lj}\kappa_{leb}^{-1}\bar{l}_3(\kappa_{leb}^{-1}\bar{l}_3 + ct + 1)$.

Note that $\{x_k\}$ converges to X^*. By (3.27), we know

$$\left\|\Sigma_{k_i,1}^{-1}\right\| = \frac{1}{\sigma_{k_i,r}} \le \left|\frac{1}{\bar{\sigma}_{k_i,r} - \kappa_{lj}\|\bar{x}_{k_i}-x_{k_i}\|}\right| \le \frac{2}{\bar{\sigma}_{k_i,r}} \tag{3.36}$$

holds for sufficiently large k_i. By (3.9),

$$\left\|\lambda_{k_i}^{-1}\Sigma_{k_i,2}\right\| = \frac{\|\Sigma_{k_i,2}\|}{\lambda_{k_i}} \le \frac{\kappa_{lj}}{\mu_{\min}\kappa_{leb}^{\delta}}\|\bar{x}_{k_i}-x_{k_i}\|^{1-\delta}. \tag{3.37}$$

It then follows from (3.33), (3.35)–(3.37), Lemma 3.5 and $1 \le \delta \le 2$ that

$$\begin{aligned}\|d_k\| &= \left\|-V_{k_i,1}\left(\Sigma_{k_i,1}^2+\lambda_{k_i}I\right)^{-1}\Sigma_{k_i,1}U_{k_i,1}^T F_k - V_{k_i,2}\left(\Sigma_{k_i,2}^2+\lambda_{k_i}I\right)^{-1}\Sigma_{k_i,2}U_{k_i,2}^T F_k\right\| \\ &\le \left\|\Sigma_{k_i,1}^{-1}\right\| \left\|U_{k_i,1}^T F_k\right\| + \left\|\lambda_{k_i}^{-1}\Sigma_{k_i,2}\right\| \left\|U_{k_i,2}^T F_k\right\| \\ &\le l_1\|\bar{x}_{k_i}-x_{k_i}\|^{k-k_i+1}, \end{aligned} \tag{3.38}$$

where $l_1 = 2\bar{\sigma}_{k_i,r}^{-1}\bar{l}_3 + \kappa_{lj}\mu_{\min}^{-1}\kappa_{leb}^{-\delta}\bar{l}_4$, and

$$\begin{aligned}
&\|F_k + G_k d_k\| \\
&\quad= \left\|\lambda_{k_i} U_{k_i,1}\left(\Sigma_{k_i,1}^2 + \lambda_{k_i} I\right)^{-1} U_{k_i,1}^T F_k + \lambda_{k_i} U_{k_i,2}\left(\Sigma_{k_i,2}^2 + \lambda_{k_i} I\right)^{-1} U_{k_i,2}^T F_k\right\| \\
&\quad\le \mu_{k_i}\|F_{k_i}\|^{\delta}\left\|\Sigma_{k_i,1}^{-2}\right\|\left\|U_{k_i,1}^T F_k\right\| + \left\|U_{k_i,2}^T F_k\right\| \\
&\quad\le l_2\|\bar{x}_{k_i} - x_{k_i}\|^{k-k_i+2},
\end{aligned} \tag{3.39}$$

where $l_2 = 4\mu_{\max}\kappa_{lf}^{\delta}\bar{\sigma}_{k_i,r}^{-2}\bar{l}_3 + \bar{l}_4$. The proof is completed. □

Based on Lemma 3.6, we have the following main result.

Theorem 3.7 *Under Assumptions 3.1–3.3, there exists a constant $l_3 > 0$ such that*

$$\|d_{k_{i+1}}\| \le l_3\|d_{k_i}\|^{s_i+1}. \tag{3.40}$$

Consequently, Algorithm 1 converges q-superlinearly to some solution of (1.1).

Proof It follows from (3.2), (3.3), Lemma 3.4 and Lemma 3.6 that

$$\begin{aligned}
\kappa_{leb}\|\bar{x}_{k_{i+1}} - x_{k_{i+1}}\| &\le \|F(x_{k_{i+1}})\| = \|F(x_{k_i+s_i-1} + d_{k_i+s_i-1})\| \\
&\le \|F_{k_i+s_i-1} + J_{k_i+s_i-1}d_{k_i+s_i-1}\| + \kappa_{lj}\|d_{k_i+s_i-1}\|^2 \\
&\le \|F_{k_i+s_i-1} + G_{k_i+s_i-1}d_{k_i+s_i-1}\| + \|(J_{k_i+s_i-1} - G_{k_i+s_i-1})d_{k_i+s_i-1}\| \\
&\quad + \kappa_{lj}\|d_{k_i+s_i-1}\|^2 \\
&\le \|F_{k_i+s_i-1} + G_{k_i+s_i-1}d_{k_i+s_i-1}\| + \sum_{j=k_i}^{k_i+s_i-2}\|d_j\|\|d_{k_i+s_i-1}\| \\
&\quad + \kappa_{lj}\|d_{k_i+s_i-1}\|^2 \\
&\le \left(l_2 + tcl_1 + \kappa_{lj}l_1^2\right)\|\bar{x}_{k_i} - x_{k_i}\|^{s_i+1}.
\end{aligned} \tag{3.41}$$

Since

$$\|\bar{x}_{k_i} - x_{k_i}\| \le \|\bar{x}_{k_{i+1}} - x_{k_i}\| \le \|\bar{x}_{k_{i+1}} - x_{k_{i+1}}\| + \sum_{j=k_i}^{k_{i+1}-1}\|d_j\|, \tag{3.42}$$

by (3.16) and (3.41),

$$\|\bar{x}_{k_i} - x_{k_i}\| \le 2\sum_{j=k_i}^{k_{i+1}-1}\|d_j\| \le 2s_i c_{s_i-1}\|d_{k_i}\| \tag{3.43}$$

holds for sufficiently large k.

Let $\bar{l}_5 = \kappa_{leb}^{-1}(l_2 + tcl_1 + \kappa_{lj}l_1^2)$. It then follows from (3.41), Lemmas 3.4 and 3.6 that

$$\begin{aligned}
\|d_{k_{i+1}}\| &\le c\|\bar{x}_{k_{i+1}} - x_{k_{i+1}}\| \le c\bar{l}_5\|\bar{x}_{k_i} - x_{k_i}\|^{s_i+1} \\
&\le c\bar{l}_5(2s_i c_{s_i-1})^{s_i+1}\|d_{k_i}\|^{s_i+1} \\
&\le c\bar{l}_5(2tc_{t-1})^{t+1}\|d_{k_i}\|^{s_i+1}.
\end{aligned} \tag{3.44}$$

Letting $l_3 = c\bar{l}_5(2tc_{t-1})^{t+1}$, we obtain (3.40).

By Lemma 3.4, we have $\|d_{k_i}\| \to 0$. Hence, there exist N and $0 < q < 1$ such that $\max\{\|d_{k_i}\|, l_3\|d_{k_i}\|\} \le q < 1$ for all $i \ge N$. Since $s_i \ge 1$, we have

$$\|d_{k_{i+1}}\| \le l_3\|d_{k_i}\|^{s_i+1} \le l_3\|d_{k_i}\|^2 \le q\|d_{k_i}\|, \qquad \forall i \ge N. \tag{3.45}$$

Then,

$$\begin{aligned}\sum_{i=N}^{\infty} \|d_{k_i}\| &\le \|d_{k_N}\| + q\|d_{k_N}\| + q^2\|d_{k_N}\| + \cdots \\ &= \frac{\|d_{k_N}\|}{1-q}.\end{aligned} \tag{3.46}$$

This implies that $\sum_{i=N}^{\infty} \|d_{k_i}\|$ converges, so does $\sum_{i=1}^{\infty} \|d_{k_i}\|$.

By (3.15) and $s_i \le t$, we have

$$\|x_{k_{i+1}} - x_{k_i}\| = \left\| \sum_{j=0}^{s_i-1} d_{k_i+j} \right\| \le tc_{t-1}\|d_{k_i}\|. \tag{3.47}$$

So, $\sum_{i=1}^{\infty} \|x_{k_{i+1}} - x_{k_i}\|$ converges, which implies that $\sum_{i=1}^{\infty}(x_{k_{i+1}} - x_{k_i})$ converges. Thus, $\{x_{k_i}\}$ converges to some $\hat{x} \in X^*$. By (3.40), we have

$$\|x_{k_{i+1}} - \hat{x}\| \le \tilde{q}\|x_{k_i} - \hat{x}\|^{s_i+1} \tag{3.48}$$

for some $\tilde{q} > 0$. Therefore, Algorithm 1 converges q-superlinearly. □

4 Numerical Results

We tested Algorithm 1 on some singular problems and compared it with the LM algorithm, the modified LM algorithm [2] as well as the multi-step LM algorithm with $m = 3$ [3]. The experiments are implemented on a laptop with an Intel Core i7-7500U CPU and 8 GB of RAM, using Matlab R2015b.

The test problems are created by modifying the nonsingular problems given by Moré, Garbow and Hillstrom in [8]. They have the same form as in [11]

$$\hat{F}(x) = F(x) - J(x^*)A\left(A^TA\right)^{-1}A^T(x - x^*), \tag{4.1}$$

where $F(x)$ is the standard nonsingular test function, x^* is its root, and $A \in R^{n\times k}$ has full column rank with $1 \le k \le n$. Obviously, $\hat{F}(x^*) = 0$ and

$$\hat{J}(x^*) = J(x^*)\left(I - A\left(A^TA\right)^{-1}A^T\right)$$

has rank $n - k$.

We set $p_0 = 0.0001$, $p_1 = 0.50$, $p_2 = 0.25$, $p_3 = 0.75$, $c_1 = 4$, $c_2 = 0.25$, $\mu_1 = 10^{-5}$, $\mu_{\min} = 10^{-8}$ and $t = 10$ for the tests. Algorithm 1 stops when $\|J_{k_i}^T F_{k_i}\| \le 10^{-5}$ or the iteration number exceeds $100 \times (n+1)$.

We chose the rank of $\hat{J}(x^*)$ to be $n - 1$ by using

$$A \in R^{n\times 1}, \qquad A^T = (1, 1, \ldots, 1).$$

The results are given in Tables 1 and 2. In the tables, x_0, $10x_0$ and $100x_0$ in the third column are starting points, where x_0 is suggested in [8]. "NF" and "NJ" represent the numbers of function

Table 1 Results on the singular problems with rank $n-1$ in small scale

Prob.	n	x_0	LM NF/NJ/NF+n*NJ	Modified LM NF/NJ/NF+n*NJ	Multi-step LM ($m=3$) NF/NJ/NF+n*NJ	Adaptive LM NF/NJ/NF+n*NJ
1	2	1	15/15/45	21/11/43	46/10/66	27/8/43
		10	17/17/51	25/13/51	31/11/53	34/9/52
		100	21/21/63	29/15/59	37/13/63	40/10/60
2	4	1	10/10/50	13/7/41	16/6/40	18/5/38
		10	13/13/65	19/10/59	22/8/54	25/7/53
		100	16/16/80	23/12/71	28/10/68	32/8/64
4	4	1	16/16/80	23/12/71	28/10/68	30/8/62
		10	19/19/95	27/14/83	34/12/82	36/9/72
		100	22/22/110	31/16/95	40/14/96	43/11/87
5	3	1	8/8/32	11/6/29	13/5/28	38/4/50
		10	8/8/32	9/5/24	28/10/58	53/6/71
		100	8/8/32	19/10/49	31/11/64	250/47/391
9	30	1	2/2/62	3/2/63	4/2/64	2/1/32
		10	14/9/284	27/14/447	37/13/427	25/5/175
		100	9/9/279	13/7/223	16/6/196	20/4/140
10	30	1	6/6/186	7/4/127	10/4/130	9/2/69
		10	7/7/217	11/6/191	13/5/163	33/5/183
		100	10/10/310	13/7/223	16/6/196	20/4/140
12	30	1	18/18/558	25/13/415	31/11/361	35/9/305
		10	20/20/620	27/14/447	34/12/394	38/10/338
		100	23/23/713	33/17/543	40/14/460	46/12/406
13	30	1	9/9/279	13/7/223	16/6/196	17/4/137
		10	14/14/434	19/10/319	25/9/295	26/7/236
		100	17/17/527	25/13/415	31/11/361	34/9/304
14	30	1	12/12/372	17/9/287	22/8/262	23/6/203
		10	18/18/558	27/14/447	31/11/361	36/10/336
		100	24/24/744	35/18/575	43/15/493	46/13/436

Table 2 Results on the singular problems with rank $n-1$ in large scale

Prob.	n	x_0	LM NF/NJ/TIME	Modified LM NF/NJ/TIME	Multi-step LM ($m=3$) NF/NJ/TIME	Adaptive LM NF/NJ/TIME
8	4000	1	9/9/12.39	13/7/10.66	16/6/9.70	18/5/6.65
9	4000	1	1/1/0.23	1/1/0.22	1/1/0.22	1/1/0.22
		10	3/3/3.36	5/3/3.85	7/3/4.08	7/1/1.67
		100	4/4/4.89	7/4/5.50	7/3/4.11	12/2/3.34
10	4000	1	7/7/100.15	11/6/102.80	13/5/91.07	13/3/68.12
13	4000	1	9/9/13.16	13/7/11.16	16/6/9.66	16/4/6.69
		10	14/14/21.55	19/10/16.72	25/9/15.64	26/7/11.47
		100	17/17/26.77	25/13/22.22	31/11/19.42	34/9/14.92

Table 2 continued

Prob.	n	x_0	LM NF/NJ/TIME	Modified LM NF/NJ/TIME	Multi-step LM ($m = 3$) NF/NJ/TIME	Adaptive LM NF/NJ/TIME
14	4000	1	12/12/18.77	17/9/15.20	22/8/14.10	23/6/10.33
		10	18/18/28.83	27/14/24.57	31/11/20.05	35/9/15.69
		100	24/24/39.00	35/18/32.10	43/15/28.10	46/13/22.39

Table 3 Results on the singular problems with rank 1 in small scale

Prob.	n	x_0	LM NF/NJ/NF+n*NJ	Modified LM NF/NJ/NF+n*NJ	Multi-step LM ($m = 3$) NF/NJ/NF+n*NJ	Adaptive LM NF/NJ/NF+n*NJ
1	2	1	15/15/45	21/11/43	25/9/43	28/7/42
		10	17/17/51	25/13/51	31/11/53	34/9/52
		100	21/21/63	29/15/59	37/13/63	40/10/60
2	4	1	10/10/50	13/7/41	16/6/40	18/5/38
		10	13/13/65	19/10/59	22/8/54	25/7/53
		100	16/16/80	23/12/71	28/10/68	32/8/64
4	4	1	16/16/80	21/11/65	28/10/68	30/8/62
		10	19/19/95	27/14/83	31/11/75	35/9/71
		100	22/22/110	31/16/95	37/13/89	42/11/86
5	3	1	15/15/60	21/11/54	25/9/52	29/8/53
		10	15/15/60	21/11/54	25/9/52	21/3/30
		100	15/15/60	21/11/54	25/9/52	28/5/43
9	30	1	2/2/62	3/2/63	4/2/64	2/1/32
		10	3/3/93	5/3/95	7/3/97	5/1/35
		100	19/10/319	37/10/337	34/12/394	17/4/137
10	30	1	3/3/93	3/2/63	4/2/64	5/2/65
		10	8/8/248	15/6/195	—	112/9/382
		100	23/14/443	727/265/867	—	36/6/216
12	30	1	18/18/558	25/13/415	31/11/361	35/9/305
		10	20/20/620	27/14/447	34/12/394	38/10/338
		100	23/23/713	33/17/543	40/14/460	46/12/406
13	30	1	9/9/279	13/7/223	16/6/196	15/4/135
		10	13/13/403	19/10/319	22/8/262	26/7/236
		100	17/17/527	23/12/383	28/10/328	32/8/272
14	30	1	11/11/341	15/8/255	19/7/229	22/6/202
		10	17/17/527	25/13/415	31/11/361	33/9/303
		100	23/23/713	33/17/543	40/14/460	43/12/403

evaluations and Jacobian evaluations, respectively. "—" represents that the algorithm fails to find a solution in $100 \times (n + 1)$ iterations. Note that, the evaluation of the Jacobian is generally n times of the function evaluation. So we also presented the values "$NF + n * NJ$" for comparisons of the total evaluations. However, if the Jacobian is sparse, this kind of value

Table 4 Results on the singular problems with rank 10 in large scale

Prob.	n	x_0	LM NF/NJ/TIME	Modified LM NF/NJ/TIME	Multi-step LM ($m = 3$) NF/NJ/TIME	Adaptive LM NF/NJ/TIME
8	4000	1	9/9/12.61	13/7/10.64	16/6/9.53	18/5/6.89
9	4000	1	1/1/0.22	1/1/0.23	1/1/0.23	1/1/0.22
		10	3/3/3.41	3/2/2.00	4/2/2.14	3/1/1.72
		100	9/9/14.42	15/8/12.72	13/5/7.92	26/4/7.41
10	4000	1	6/6/85.63	7/4/64.68	10/4/71.46	9/3/57.73
13	4000	1	8/8/11.78	11/6/9.16	16/6/9.67	15/4/6.88
		10	13/13/20.24	19/10/16.21	22/8/13.50	26/7/12.06
		100	17/17/26.88	23/12/19.81	28/10/17.22	33/9/14.25
14	4000	1	11/11/17.27	15/8/13.16	19/7/11.81	21/6/9.16
		10	17/17/27.64	23/12/20.28	31/11/19.78	33/9/16.03
		100	23/23/37.60	31/16/27.79	40/14/25.47	43/12/21.28

does not mean much. For large scale problems, we presented the running time of the problem instead of the total evaluations.

We also chose the rank of $\hat{J}(x^*)$ to be 1 and 10, where $A \in R^{n\times(n-1)}$ and $A \in R^{n\times(n-10)}$ are generated randomly. The results are given in Tables 3 and 4, respectively.

From the tables, we can see that the adaptive LM algorithm usually takes the least Jacobian calculations as well as the total calculations to find a solution of small scale nonlinear equations, while it usually takes the least time to find a solution of large scale nonlinear equations, no matter the rank of the problem is. The reason is that it makes full use of the available evaluated Jacobians and matrix factorizations to compute approximate LM steps. Since it takes much less efforts to compute approximate LM steps, the overall cost of the adaptive LM algorithm is usually much less.

5 Conclusion

We proposed an adaptive multi-step Levenberg–Marquardt algorithm for nonlinear equations. It uses the latest evaluated Jacobian to calculate an approximate LM step, if the ratio of the actual reduction to the predicted reduction of the merit function at the iterate is good. Under the local error bound condition, which is weaker than the nonsingularity condition, the adaptive LM algorithm converges superlinearly to some solution of (1.1). Compared to the LM algorithm, the Jacobian calculations, the total calculations, as well as the running time of the adaptive multi-step LM algorithm are significantly reduced.

References

1. Behling, R., Iusem, A.: The effect of calmness on the solution set of systems of nonlinear equations. Math. Program. **137**, 155–165 (2013)
2. Fan, J.Y.: The modified Levenberg–Marquardt method for nonlinear equations with cubic convergence. Math. Comput. **81**, 447–466 (2012)
3. Fan, J.Y.: A Shamanskii-like Levenberg–Marquardt method for nonlinear equations. Comput. Optim. Appl. **56**, 63–80 (2013)

4. Fan, J.Y., Yuan, Y.X.: On the quadratic convergence of the Levenberg–Marquardt method without nonsingularity assumption. Computing **74**, 23–39 (2005)
5. Kelley, C.T.: Solving Nonlinear Equations with Newton's Method. Fundamentals of Algorithms. SIAM, Philadelphia (2003)
6. Levenberg, K.: A method for the solution of certain nonlinear problems in least squares. Q. Appl. Math. **2**, 164–166 (1944)
7. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear inequalities. SIAM J. Appl. Math. **11**, 431–441 (1963)
8. Moré, J.J., Garbow, B.S., Hillstrom, K.E.: Testing unconstrained optimization software. ACM Trans. Math. Softw. (TOMS) **7**, 17–41 (1981)
9. Moré, J.J.: The Levenberg–Marquardt Algorithm: Implementation and Theory. In: Watson, G.A. (ed.) Lecture Notes in Mathematics 630: Numerical Analysis, pp. 105–116. Springer, Berlin (1978)
10. Powell, M.J.D.: Convergence properties of a class of minimization algorithms. Nonlinear Program. **2**, 1–27 (1975)
11. Schnabel, R.B., Frank, P.D.: Tensor methods for nonlinear equations. SIAM J. Numer. Anal. **21**, 815–843 (1984)
12. Stewart, G.W., Sun, J.-G.: Matrix Perturbation Theory, (Computer Science and Scientific Computing). Academic Press, Boston (1990)
13. Yamashita, N., Fukushima, M.: On the rate of convergence of the Levenberg–Marquardt method. Comput. Suppl. **15**, 237–249 (2001)
14. Yuan, Y.X.: Recent advances in trust region algorithms. Math. Program. Ser. B **151**, 249–281 (2015)