J Sci Comput (2018) 74:1146–1162
https://doi.org/10.1007/s10915-017-0488-6

# On a New Updating Rule of the Levenberg–Marquardt Parameter

**Ruixue Zhao**[1] · **Jinyan Fan**[1]

Received: 7 June 2016 / Revised: 21 January 2017 / Accepted: 16 June 2017 / Published online: 23 June 2017


**Abstract** A new Levenberg–Marquardt (LM) algorithm is proposed for nonlinear equations, where the iterate is updated according to the ratio of the actual reduction to the predicted reduction as usual, but the update of the LM parameter is no longer just based on that ratio. When the iteration is unsuccessful, the LM parameter is increased; but when the iteration is successful, it is updated based on the value of the gradient norm of the merit function. The algorithm converges globally under certain conditions. It also converges quadratically under the local error bound condition, which does not require the nonsingularity of the Jacobian at the solution.

**Keywords** Levenberg–Marquardt method · Trust region method · Nonlinear equations · Local error bound · Quadratic convergence

## 1 Introduction

We consider the system of nonlinear equations

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

where $F(x) : R^n \to R^n$ is continuously differentiable.

The Levenberg–Marquardt method (LM) is one of the most well-known iterative methods for nonlinear equations [5,6,15]. At the $k$-th iteration, it computes the trial step

$$d_k = -(J_k^T J_k + \lambda_k I)^{-1} J_k^T F_k, \tag{1.2}$$

Jinyan Fan is partially supported by the NSFC Grant 11571234.

✉ Jinyan Fan
jyfan@sjtu.edu.cn

1 School of Mathematical Sciences, and MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, China

where $F_k = F(x_k)$, $J_k = J(x_k)$ is the Jacobian at $x_k$, and $\lambda_k$ is the LM parameter introduced to overcome the difficulties caused by the singularity or near singularity of $J_k$.

Let

$$\min_{x \in R^n} \quad \phi(x) := \|F(x)\|^2 \tag{1.3}$$

be the merit function of (1.1). Define the actual reduction of the merit function as

$$Ared_k = \|F_k\|^2 - \|F(x_k + d_k)\|^2,$$

the predicted reduction as

$$Pred_k = \|F_k\|^2 - \|F_k + J_k d_k\|^2,$$

and the ratio of the actual reduction to the predicted reduction

$$r_k = \frac{Ared_k}{Pred_k}.$$

In classical LM methods, one sets

$$x_{k+1} = \begin{cases} x_k + d_k, & \text{if } r_k \geq p_0, \\ x_k, & \text{otherwise}, \end{cases} \tag{1.4}$$

where $p_0 \geq 0$ is a constant, and updates the LM parameter as

$$\lambda_{k+1} = \begin{cases} c_0 \lambda_k, & \text{if } r_k < p_1, \\ \lambda_k, & \text{if } r_k \in [p_1, p_2], \\ c_1 \lambda_k, & \text{if } r_k > p_2, \end{cases} \tag{1.5}$$

where $p_0 < p_1 < p_2 < 1, 0 < c_1 < 1 < c_0$ are positive constants (cf. [7,9,13,16,17]).

It was shown in [14] that, if the LM parameter is chosen as $\lambda_k = \|F_k\|^2$, then the LM method converges quadratically under the local error bound condition, which is weaker than the nonsingularity of the Jacobian at the solution. It was further proved in [4] that the LM method converges quadratically for all $\lambda_k = \|F_k\|^\delta (\delta \in [1, 2])$ under the local error bound condition. In [1], Fan chose

$$\lambda_k = \mu_k \|F_k\|, \tag{1.6}$$

and updated $\mu_k$ according to the ratio $r_k$ as follows:

$$\mu_{k+1} = \begin{cases} c_0 \mu_k, & \text{if } r_k < p_1, \\ \mu_k, & \text{if } r_k \in [p_1, p_2], \\ \max\{c_1 \mu_k, m\}, & \text{if } r_k > p_2, \end{cases} \tag{1.7}$$

where $m > 0$ is a small constant to prevent the LM parameter from being too small.

Recently, Zhao and Fan [18] took the LM parameter as

$$\lambda_k = \mu_k \|J_k^T F_k\|, \tag{1.8}$$

where the update of $\mu_k$ is no longer just based on the ratio $r_k$. When the iteration is unsuccessful (i.e., $r_k < p_0$), $\mu_k$ is increased; but when the iteration is successful (i.e., $r_k \geq p_0$), $\mu_{k+1}$ is updated as

$$\mu_{k+1} = \begin{cases} c_0 \mu_k, & \text{if } \|J_k^T F_k\| < \dfrac{p_1}{\mu_k}, \\ \mu_k, & \text{if } \|J_k^T F_k\| \in [\dfrac{p_1}{\mu_k}, \dfrac{p_2}{\mu_k}], \\ \max\{c_1 \mu_k, m\}, & \text{if } \|J_k^T F_k\| > \dfrac{p_2}{\mu_k}. \end{cases} \tag{1.9}$$

It was shown that the global complexity bound of the above LM algorithm is $O(\varepsilon^{-2})$, that is, it takes at most $O(\varepsilon^{-2})$ iterations to derive the norm of the gradient of the merit function below the desired accuracy $\varepsilon$.

The logic behind the updating rule (1.9) follows from the fact that the LM step is actually the solution of the trust region subproblem

$$\begin{aligned} &\min_{d\in R^n} \|F_k + J_k d\|^2 \\ &s.t. \ \|d\| \leq \Delta_k := \|d_k\|. \end{aligned} \tag{1.10}$$

So, the step size computed by solving (1.10) is proportional to the norm of the model gradient $\|J_k^T F_k\|$. Hence, the trust region, a magnitude of the inverse of $\mu_k$, should also be of comparable size.

In this paper, we present a new LM algorithm for (1.1), where the LM parameter is computed as

$$\lambda_k = \mu_k \|F_k\|^2. \tag{1.11}$$

We update the iterate $x_k$ according to the ratio $r_k$ as classical LM algorithms. When the iteration is unsuccessful, we increase $\mu_k$; otherwise, we update $\mu_{k+1}$ by (1.9). We show that the new LM algorithm preserves the global convergence of classical LM algorithms. We also prove that the algorithm converges quadratically under the local error bound condition.

The paper is organized as follows. In Sect. 2, we present the new LM algorithm for (1.1). The global convergence of the algorithm is also proved. In Sect. 3, we study 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 The LM Algorithm and Global Convergence

In this section, we first give the new LM algorithm, then show that the algorithm converges globally under certain conditions.

The LM algorithm is presented as follows.

**Algorithm 2.1** (*A Levenberg–Marquardt algorithm for nonlinear equations*)

Step 1. Given $x_0 \in R^n$, $\mu_0 > m > 0$, $0 < p_0 < p_1 < p_2 < 1$, $c_0 > 1$, $0 < c_1 < 1$, $\varepsilon \geq 0$, $k := 0$.

Step 2. If $\|J_k^T F_k\| \leq \varepsilon$, then stop. Otherwise, solve

$$(J_k^T J_k + \lambda_k I)d = -J_k^T F_k \text{ with } \lambda_k = \mu_k \|F_k\|^2 \tag{2.1}$$

to obtain $d_k$.

Step 3. Compute $r_k = \frac{Ared_k}{Pred_k}$.
If $r_k \geq p_0$, set $x_{k+1} = x_k + d_k$ and compute $\mu_{k+1}$ by (1.9);
Otherwise, set $x_{k+1} = x_k$ and compute $\mu_{k+1} = c_0 \mu_k$.
Set $k := k + 1$ and go to step 2.

To study the global convergence of Algorithm 2.1, we make the following assumption.

**Assumption 2.1** $F(x)$ is continuously differentiable, both $F(x)$ and its Jacobian $J(x)$ is Lipschitz continuous, i.e., there exist positive constants $L_1$ and $L_2$ such that

$$\|J(x) - J(y)\| \leq L_1 \|y - x\|, \quad \forall x, y \in R^n \tag{2.2}$$

and

$$\|F(x) - F(y)\| \leq L_2\|y - x\|, \quad \forall x, y \in R^n. \tag{2.3}$$

Due to the result given by Powell [10], we have the following lemma.

**Lemma 2.1** *The predicted reduction satisfies*

$$\|F_k\|^2 - \|F_k + J_k d_k\|^2 \geq \|J_k^T F_k\| \min\left\{\|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\} \tag{2.4}$$

*for all $k$.*

Lemma 2.1 implies that the predicted reduction is always nonnegative.

In the following, we first prove the weak global convergence of Algorithm 2.1, that is, at least one accumulation point of the sequence generated by Algorithm 2.1 is a stationary point of the merit function $\phi(x)$.

**Theorem 2.1** *Under Assumption 2.1, Algorithm 2.1 terminates in finite iterations or satisfies*

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

*Proof* We prove by contradiction. Suppose that there exists a constant $\tau > 0$ such that

$$\|J_k^T F_k\| \geq \tau, \quad \forall k. \tag{2.5}$$

Define the index set of successful iterations:

$$S = \{k : r_k \geq p_0\}.$$

We discuss in two cases.

Case I. $S$ is infinite. Since $\|F(x)\|$ is nonincreasing and bounded below, it follows from (2.3) and (2.4) that

$$\begin{aligned} +\infty &> \sum_{k\in S}(\|F_k\|^2 - \|F_{k+1}\|^2) \\ &\geq \sum_{k\in S} p_0(\|F_k\|^2 - \|F_k + J_k d_k\|^2) \\ &\geq \sum_{k\in S} p_0\|J_k^T F_k\| \min\left\{\|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\} \\ &\geq \sum_{k\in S} p_0\tau \min\left\{\|d_k\|, \frac{\tau}{L_2^2}\right\}. \end{aligned} \tag{2.6}$$

So,

$$\lim_{k\in S, k\to\infty} d_k = 0. \tag{2.7}$$

Note that $d_k = 0$ for $k \notin S$, we have

$$\lim_{k\to\infty} d_k = 0. \tag{2.8}$$

Since $\|J_k\| \leq L_2$ and $\|F_k\| \leq \|F_0\|$, by (2.1), we have

$$\mu_k \to +\infty. \tag{2.9}$$

On the other hand, it follows from (2.3) and (2.4) that

$$\begin{aligned}|r_k - 1| &= \left|\frac{Ared_k - Pred_k}{Pred_k}\right| \\ &= \frac{|\|F(x_k + d_k)\|^2 - \|F_k + J_k d_k\|^2|}{Pred_k} \\ &\le \frac{\|F_k + J_k d_k\| O(\|d_k\|^2) + O(\|d_k\|^4)}{\tau \min\{\|d_k\|, \tau/L_2^2\}} \\ &\to 0. \end{aligned} \tag{2.10}$$

So, $r_k \to 1$. Thus, $\mu_k$ is updated by (1.9) and $\|J_k^T F_k\| > p_2/\mu_k$ for all sufficiently large $k$. Hence, $\mu_k = \max\{c_1\mu_k, m\}$ for all large $k$. Note that $0 < c_1 < 1$, there exists a positive constant $\tilde{c}$ such that

$$\mu_k < \tilde{c}$$

for all large $k$. This is a contradiction to (2.9).

Case II. $S$ is finite. Then there exists a $\tilde{k}$ such that

$$r_k < p_0, \quad k \ge \tilde{k}. \tag{2.11}$$

According to the updating rule of $x_k$ in Algorithm 2.1, we have $d_k \to 0$. By the same arguments as (2.10), we get $r_k \to 1$, which contradicts (2.11). The proof is completed. □

Based on Theorem 2.1, we can further prove the strong global convergence of Algorithm 2.1, that is, all limit points of the sequence generated by Algorithm 2.1 are stationary points of the merit function $\phi(x)$. We first give an auxiliary result (cf. [3, Lemma 2.7]).

**Lemma 2.2** *Let $b, a_1, \ldots, a_N > 0$. Then,*

$$\sum_{j=1}^{N} \min\{a_j, b\} \ge \min\left\{\sum_{j=1}^{N} a_j, b\right\}. \tag{2.12}$$

**Theorem 2.2** *Under Assumption 2.1, Algorithm 2.1 terminates in finite iterations or satisfies*

$$\lim_{k\to\infty} \|J_k^T F_k\| = 0. \tag{2.13}$$

*Proof* Suppose by contradiction that there exists $\tau > 0$ such that the set

$$\Omega = \{k : \|J_k^T F_k\| \ge \tau\} \tag{2.14}$$

is infinite. Given $k \in \Omega$, consider the first index $l_k > k$ such that $\|J_{l_k}^T F_{l_k}\| \le \frac{\tau}{2}$. The existence of such $l_k$ is guaranteed by Theorem 2.1. By (2.2), (2.3) and $\|F_k\| \le \|F_0\|$,

$$\begin{aligned}\frac{\tau}{2} &\le \|J_k^T F_k\| - \|J_{l_k}^T F_{l_k}\| \le \|J_k^T F_k - J_{l_k}^T F_{l_k}\| \\ &\le \|J_k^T F_k - J_{l_k}^T F_k\| + \|J_{l_k}^T F_k - J_{l_k}^T F_{l_k}\| \le (L_1\|F_0\| + L_2^2)\|x_k - x_{l_k}\|,\end{aligned}$$

which yields

$$\|x_k - x_{l_k}\| \ge \frac{\tau}{2(L_1\|F_0\| + L_2^2)}.$$

Define the set

$$S_k = \{j : k \le j < l_k, x_{j+1} \ne x_j\}.$$

Then,

$$\frac{\tau}{2(L_1\|F_0\|+L_2^2)} \le \|x_k - x_{l_k}\| \le \sum_{j\in S_k}\|x_j - x_{j+1}\| \le \sum_{j\in S_k}\|d_j\|. \tag{2.15}$$

It now follows from (2.4), (2.15) and Lemma 2.2 that, for all $k \in \Omega$,

$$\begin{aligned}
\|F_k\|^2 - \|F_{l_k}\|^2 &= \sum_{j\in S_k}(\|F_j\|^2 - \|F_{j+1}\|^2)\\
&\ge \sum_{j\in S_k} p_0\|J_j^T F_j\| \min\left\{\|d_j\|, \frac{\|J_j^T F_j\|}{\|J_j^T J_j\|}\right\}\\
&\ge \sum_{j\in S_k}\frac{p_0\tau}{2}\min\left\{\|d_j\|, \frac{\tau}{2L_2^2}\right\}\\
&\ge \frac{p_0\tau}{2}\min\left\{\sum_{j\in S_k}\|d_j\|, \frac{\tau}{2L_2^2}\right\}\\
&\ge \frac{p_0\tau}{2}\min\left\{\frac{\tau}{2(L_1\|F_0\|+L_2^2)}, \frac{\tau}{2L_2^2}\right\}\\
&= \frac{p_0\tau^2}{4(L_1\|F_0\|+L_2^2)}\\
&> 0.
\end{aligned} \tag{2.16}$$

However, since $\{\|F_k\|^2\}$ is nonincreasing and bounded below, $\|F_k\|^2 - \|F_{l_k}\|^2 \to 0$. This contradicts (2.16). So, the set $\Omega$ defined by (2.14) is finite. Therefore, (2.13) holds true. The proof is completed. □

## 3 Local Convergence

We assume that the sequence $\{x_k\}$ generated by Algorithm 2.1 converges to the solution set $X^*$ of (1.1) and lies in some neighbourhood of $x^* \in X^*$. We first give some important properties of the algorithm, then show that the algorithm converges quadratically under the local error bound condition.

We make the following assumption.

**Assumption 3.1** (a) $F(x)$ is continuously differentiable, and $\|F(x)\|$ provides a local error bound on some neighbourhood of $x^* \in X^*$, i.e., there exist positive constants $c > 0$ and $b_1 < 1$ such that

$$\|F(x)\| \ge c\,\mathrm{dist}(x, X^*), \qquad \forall x \in N(x^*, b_1) = \{x : \|x - x^*\| \le b_1\}. \tag{3.1}$$

(b) The Jacobian $J(x)$ is Lipschitz continuous on $N(x^*, b_1)$, i.e., there exists a positive constant $L_1$ such that

$$\|J(y) - J(x)\| \le L_1\|y - x\|, \quad \forall x, y \in N(x^*, b_1). \tag{3.2}$$

Note that, if $J(x)$ is nonsingular at a solution of (1.1), then it is an isolated solution, so $\|F(x)\|$ provides a local error bound on its neighborhood. However, the converse is not necessarily true. Please see examples in [14]. Thus, the local error bound condition is weaker than the nonsingularity.

By (3.2), we have

$$\|F(y) - F(x) - J(x)(y - x)\| \le L_1\|y - x\|^2, \quad \forall x, y \in N(x^*, b_1). \tag{3.3}$$

Moreover, there exists a constant $L_2 > 0$ such that

$$\|F(y) - F(x)\| \le L_2\|y - x\|, \quad \forall x, y \in N(x^*, b_1). \tag{3.4}$$

Throughout the paper, we denote by $\bar{x}_k$ the vector in $X^*$ that satisfies

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

### 3.1 Some Properties

In the following, we first show the relationship between the length of the trial step $d_k$ and the distance from $x_k$ to the solution set.

**Lemma 3.1** *Under Assumption 3.1, if $x_k \in N(x^*, b_1/2)$, then*

$$\|d_k\| \le c_2\|\bar{x}_k - x_k\| \tag{3.5}$$

*holds for all sufficiently large k, where $c_2 = \sqrt{L_1^2 c^{-2} m^{-1} + 1}$ is a positive constant.*

*Proof* Since $x_k \in N(x^*, b_1/2)$, we have

$$\|\bar{x}_k - x^*\| \le \|\bar{x}_k - x_k\| + \|x_k - x^*\| \le 2\|x_k - x^*\| \le b_1.$$

So, $\bar{x}_k \in N(x^*, b_1)$. Thus, it follows from (1.9) and (3.1) that the LM parameter $\lambda_k$ satisfies

$$\lambda_k = \mu_k\|F_k\|^2 \ge c^2 m\|\bar{x}_k - x_k\|^2. \tag{3.6}$$

Note that $d_k$ is also a minimizer of

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

by (3.3) and (3.6), we have

$$\begin{aligned}\|d_k\|^2 &\le \frac{\varphi_k(d_k)}{\lambda_k} \\ &\le \frac{\varphi_k(\bar{x}_k - x_k)}{\lambda_k} \\ &= \frac{\|F_k + J_k(\bar{x}_k - x_k)\|^2}{\lambda_k} + \|\bar{x}_k - x_k\|^2 \\ &\le \frac{L_1^2\|\bar{x}_k - x_k\|^4}{\lambda_k} + \|\bar{x}_k - x_k\|^2 \\ &\le (L_1^2 c^{-2} m^{-1} + 1)\|\bar{x}_k - x_k\|^2.\end{aligned}$$

So, we obtain (3.5). □

Next we show that the gradient of the merit function also provides a local error bound on some neighbourhood of $x^* \in X^*$.

**Lemma 3.2** *Under Assumption 3.1, if $x_k \in N(x^*, b_1/2)$, then there exists a constant $c_3 > 0$ such that*

$$\|J_k^T F_k\| \geq c_3 \|\bar{x}_k - x_k\| \tag{3.7}$$

*holds for all sufficiently large $k$.*

*Proof* It follows from (3.3) that

$$\|F_k + J_k(\bar{x}_k - x_k)\| \leq L_1 \|\bar{x}_k - x_k\|^2.$$

Thus,

$$\|F_k\|^2 + 2(\bar{x}_k - x_k)^T J_k^T F_k + (\bar{x}_k - x_k)^T J_k^T J_k(\bar{x}_k - x_k) \leq L_1^2 \|\bar{x}_k - x_k\|^4.$$

So,

$$\|F_k\|^2 + 2(\bar{x}_k - x_k)^T J_k^T F_k \leq L_1^2 \|\bar{x}_k - x_k\|^4.$$

By (3.1),

$$c^2 \|\bar{x}_k - x_k\|^2 - L_1^2 \|\bar{x}_k - x_k\|^4 \leq 2\|\bar{x}_k - x_k\| \|J_k^T F_k\|.$$

Hence, (3.7) holds for sufficiently large $k$. The proof is completed. □

**Lemma 3.3** *Under Assumption 3.1, if $x_k \in N(x^*, b_1/2)$, then there exists a positive integer $K$ such that*

$$r_k \geq p_0, \quad \forall k \geq K.$$

*That is, $\mu_k$ is updated by (1.9) when $k \geq K$.*

*Proof* It follows form (3.4), (3.5) and (3.7) that

$$\begin{aligned} Pred_k &\geq \|J_k^T F_k\| \min\left\{\|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\} \\ &\geq c_3 \|\bar{x}_k - x_k\| \min\{\|d_k\|, \frac{c_2^{-1} c_3}{L_2^2} \|d_k\|\} \\ &= \|d_k\| O(\|\bar{x}_k - x_k\|). \end{aligned}$$

This, together with (3.3), (3.4) and $\|F_k + J_k d_k\| \leq \|F_k\|$, gives

$$\begin{aligned} |r_k - 1| &= |\frac{Ared_k - Pred_k}{Pred_k}| \\ &\leq \frac{\|F_k + J_k d_k\| O(\|d_k\|^2) + O(\|d_k\|^4)}{Pred_k} \\ &\leq \frac{O(\|\bar{x}_k - x_k\|) O(\|d_k\|^2) + O(\|d_k\|^4)}{\|d_k\| O(\|\bar{x}_k - x_k\|)} \\ &= O(\|d_k\|) \\ &\to 0. \end{aligned}$$

So, $r_k \to 1$. Therefore, we obtain the result. □

Let

$$C_1 = \max\{p_2, c_1^{-1} m L_2 \|F_0\|\}, \tag{3.8}$$

$$c_4 = L_2^2 + L_1 \|F_0\| \tag{3.9}$$

be two positive constants.

**Lemma 3.4** *Under Assumption 3.1 and* $c_1 \leq (1 + c_4 c_2 c_3^{-1})^{-1}$*, if* $k \geq K$ *and* $\mu_k \| J_k^T F_k \| > C_1$*, then*

$$\mu_{k+1} \| J_{k+1}^T F_{k+1} \| \leq \mu_k \| J_k^T F_k \|. \tag{3.10}$$

*Proof* By (3.2) and (3.4),

$$\begin{aligned} | \| J_{k+1}^T F_{k+1} \| - \| J_k^T F_k \| | \leq & | \| J_{k+1}^T F_{k+1} \| - \| J_{k+1}^T F_k \| | + | \| J_{k+1}^T F_k \| - \| J_k^T F_k \| | \\ \leq & \| J_{k+1} \| \| F_{k+1} - F_k \| + \| F_k \| \| J_{k+1} - J_k \| \\ \leq & (L_2^2 + L_1 \| F_0 \|) \| d_k \| \\ = & c_4 \| d_k \|. \end{aligned}$$

It then follows from Lemmas 3.1 and 3.2 that

$$\| J_{k+1}^T F_{k+1} \| \leq \| J_k^T F_k \| + c_4 \| d_k \| \leq (1 + c_4 c_2 c_3^{-1}) \| J_k^T F_k \|. \tag{3.11}$$

Since $\mu_k \| J_k^T F_k \| > C_1$, by (3.4) and $\| F_k \| \leq \| F_0 \|$, we have

$$\mu_k > \frac{p_2}{\| J_k^T F_k \|}, \quad \mu_k \| J_k^T F_k \| \geq \frac{m L_2}{c_1} \| F_0 \| \geq \frac{m}{c_1} \| J_k^T F_k \|.$$

So, $\mu_k \geq \frac{m}{c_1}$. It then follows from $k \geq K$, Lemma 3.3 and the updating rule (1.9) that

$$\mu_{k+1} = c_1 \mu_k.$$

By (3.11) and $c_1 \leq (1 + c_4 c_2 c_3^{-1})^{-1}$, we have

$$\begin{aligned} \mu_{k+1} \| J_{k+1}^T F_{k+1} \| = & c_1 \mu_k \| J_{k+1}^T F_{k+1} \| \\ \leq & c_1 (1 + c_4 c_2 c_3^{-1}) \mu_k \| J_k^T F_k \| \\ \leq & \mu_k \| J_k^T F_k \|. \end{aligned}$$

The proof is completed. □

Let

$$C_2 = \max\{ \mu_K \| J_K^T F_K \|, c_0 (1 + c_4 c_2 c_3^{-1}) C_1 \}$$

be a positive constant.

The next lemma shows that $\mu_k \| J_k^T F_k \|$ is upper bounded.

**Lemma 3.5** *Under conditions of Lemma 3.4,*

$$\mu_k \| J_k^T F_k \| \leq C_2, \quad \forall k \geq K. \tag{3.12}$$

*Proof* We discuss in two cases.

**Case 1** $\mu_K \| J_K^T F_K \| \leq c_0 (1 + c_4 c_2 c_3^{-1}) C_1$. Then, we must have

$$\mu_{K+1} \| J_{K+1}^T F_{K+1} \| \leq c_0 (1 + c_4 c_2 c_3^{-1}) C_1. \tag{3.13}$$

Otherwise, suppose

$$\mu_{K+1} \| J_{K+1}^T F_{K+1} \| > c_0 (1 + c_4 c_2 c_3^{-1}) C_1. \tag{3.14}$$

It follows from (3.11) and $\mu_{K+1} \leq c_0 \mu_K$ that

$$(1 + c_4 c_2 c_3^{-1}) C_1 < \mu_K \| J_{K+1}^T F_{K+1} \| \leq (1 + c_4 c_2 c_3^{-1}) \mu_K \| J_K^T F_K \|. \tag{3.15}$$

This gives

$$\mu_K \|J_K^T F_K\| > C_1.$$

By Lemma 3.4, we obtain

$$\mu_{K+1} \|J_{K+1}^T F_{K+1}\| \le \mu_K \|J_K^T F_K\| \le c_0(1 + c_4 c_2 c_3^{-1})C_1.$$

This is a contradiction to (3.14). So (3.13) holds true.

By induction, we can obtain

$$\mu_k \|J_k^T F_k\| \le c_0(1 + c_4 c_2 c_3^{-1})C_1, \quad \forall k \ge K. \tag{3.16}$$

**Case 2** $\mu_K \|J_K^T F_K\| > c_0(1 + c_4 c_2 c_3^{-1})C_1$. Note that $c_0 > 1$, we have

$$\mu_K \|J_K^T F_K\| > C_1.$$

So, by Lemma 3.4,

$$\mu_{K+1} \|J_{K+1}^T F_{K+1}\| \le \mu_K \|J_K^T F_K\|. \tag{3.17}$$

If $\mu_{K+1} \|J_{K+1}^T F_{K+1}\| > c_0(1 + c_4 c_2 c_3^{-1})C_1$, then by Lemma 3.4 and (3.17),

$$\mu_{K+2} \|J_{K+2}^T F_{K+2}\| \le \mu_{K+1} \|J_{K+1}^T F_{K+1}\|. \tag{3.18}$$

Otherwise, if $\mu_{K+1} \|J_{K+1}^T F_{K+1}\| \le c_0(1 + c_4 c_2 c_3^{-1})C_1$, then by the same arguments as in case 1, we have

$$\mu_{K+2} \|J_{K+2}^T F_{K+2}\| \le c_0(1 + c_4 c_2 c_3^{-1})C_1. \tag{3.19}$$

In view of (3.17)–(3.19), we obtain

$$\begin{aligned} \mu_{K+2} \|J_{K+2}^T F_{K+2}\| &\le \max\{\mu_{K+1} \|J_{K+1}^T F_{K+1}\|, c_0(1 + c_4 c_2 c_3^{-1})C_1\} \\ &\le \max\{\mu_K \|J_K^T F_K\|, c_0(1 + c_4 c_2 c_3^{-1})C_1\}. \end{aligned} \tag{3.20}$$

By induction, we can prove that, for all $k > K$,

$$\begin{aligned} \mu_k \|J_k^T F_k\| &\le \max\{\mu_{k-1} \|J_{k-1}^T F_{k-1}\|, c_0(1 + c_4 c_2 c_3^{-1})C_1\} \\ &\le \cdots \\ &\le \max\{\mu_K \|J_K^T F_K\|, c_0(1 + c_4 c_2 c_3^{-1})C_1\} \\ &= C_2. \end{aligned}$$

The proof is completed. □

Let

$$C_3 = c_3^{-1} L_2 C_2.$$

The following lemma shows that $\mu_k \|F_k\|$ is bounded by $C_3$.

**Lemma 3.6** *Under conditions of Lemma 3.4,*

$$\mu_k \|F_k\| \le C_3 \tag{3.21}$$

*holds for all sufficiently large $k$.*

*Proof* It follows from (3.4) that

$$\|F_k\| \le L_2\|\bar{x}_k - x_k\|.$$

This, together with (3.7), gives

$$\|F_k\| \le c_3^{-1} L_2\|J_k^T F_k\|.$$

Thus, by (3.12), we obtain (3.21). The proof is completed. □

### 3.2 Quadratic Convergence

Based on the above lemmas, we study the quadratic convergence of Algorithm 2.1 under the local error bound condition, by using the singular value decomposition (SVD) technique.

Suppose the SVD of $J(\bar{x}_k)$ is

$$\begin{aligned}\bar{J}_k &= \bar{U}_k \bar{\Sigma}_k \bar{V}_k^T \\ &= (\bar{U}_{k,1}, \bar{U}_{k,2}) \begin{pmatrix} \bar{\Sigma}_{k,1} & \\ & 0 \end{pmatrix} \begin{pmatrix} \bar{V}_{k,1}^T \\ \bar{V}_{k,2}^T \end{pmatrix} \\ &= \bar{U}_{k,1} \bar{\Sigma}_{k,1} \bar{V}_{k,1}^T,\end{aligned}$$

where $\bar{\Sigma}_{k,1} = \mathrm{diag}(\bar{\sigma}_{k,1}, \ldots, \bar{\sigma}_{k,r})$ with $\bar{\sigma}_{k,1} \ge \bar{\sigma}_{k,2} \ge \cdots \ge \bar{\sigma}_{k,r} > 0$, and the correspondingly SVD of $J_k$ is

$$\begin{aligned}J_k &= U_k \Sigma_k V_k^T \\ &= (U_{k,1}, U_{k,2}) \begin{pmatrix} \Sigma_{k,1} & \\ & \Sigma_{k,2} \end{pmatrix} \begin{pmatrix} V_{k,1}^T \\ V_{k,2}^T \end{pmatrix} \\ &= U_{k,1} \Sigma_{k,1} V_{k,1}^T + U_{k,2} \Sigma_{k,2} V_{k,2}^T,\end{aligned}$$

where $\Sigma_{k,1} = \mathrm{diag}(\sigma_{k,1}, \ldots, \sigma_{k,r})$ with $\sigma_{k,1} \ge \cdots \ge \sigma_{k,r} > 0$, and $\Sigma_{k,2} = \mathrm{diag}(\sigma_{k,r+1}, \ldots, \sigma_{k,n})$ with $\sigma_{k,r} \ge \cdots \ge \sigma_{k,n} \ge 0$. In the following, if the context is clear, we neglect the subscription $k$ in $\Sigma_{k,i}$ and $U_{k,i}$, $V_{k,i} (i = 1, 2)$, and write $J_k$ as

$$J_k = U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T.$$

By the theory of matrix perturbation [12] and the Lipschitzness of $J_k$,

$$\|\mathrm{diag}(\Sigma_1 - \bar{\Sigma}_1, \Sigma_2)\| \le \|J_k - \bar{J}_k\| \le L_1\|\bar{x}_k - x_k\|.$$

So,

$$\|\Sigma_1 - \bar{\Sigma}_1\| \le L_1\|\bar{x}_k - x_k\| \quad \text{and} \quad \|\Sigma_2\| \le L_1\|\bar{x}_k - x_k\|. \tag{3.22}$$

Since $\{x_k\}$ converges to the solution set $X^*$, we assume that $L_1\|\bar{x}_k - x_k\| \le \bar{\sigma}_r/2$ holds for all sufficiently large $k$. Then, it follows from (3.22) that

$$\|\Sigma_1^{-1}\| \le \frac{1}{\bar{\sigma}_r - L_1\|\bar{x}_k - x_k\|} \le \frac{2}{\bar{\sigma}_r}. \tag{3.23}$$

**Lemma 3.7** *Under Assumption 3.1, if $x_k \in N(x^*, b_1/2)$, then we have*

*(a)* $\|U_1 U_1^T F_k\| \le L_2\|\bar{x}_k - x_k\|$;
*(b)* $\|U_2 U_2^T F_k\| \le 2L_1\|\bar{x}_k - x_k\|^2$;

*where $L_1, L_2$ are given in (3.2) and (3.4) respectively.*

*Proof* (a) follows from (3.4) directly.
Denote $F(\bar{x}_k)$ by $\bar{F}_k$. By (3.3) and (3.22),

$$\begin{aligned}\|U_2U_2^TF_k\| =&\|U_2U_2^T(\bar{F}_k - F_k)\|\\ \le&\|U_2U_2^TJ_k(\bar{x}_k - x_k)\| + L_1\|U_2U_2^T\|\|\bar{x}_k - x_k\|^2\\ \le&\|U_2U_2^T(U_1\Sigma_1V_1^T + U_2\Sigma_2V_2^T)\|\|\bar{x}_k - x_k\| + L_1\|\bar{x}_k - x_k\|^2\\ \le&\|\Sigma_2\|\|\bar{x}_k - x_k\| + L_1\|\bar{x}_k - x_k\|^2\\ \le&2L_1\|\bar{x}_k - x_k\|^2.\end{aligned}$$

The proof is completed. □

Now we can give the main result of this section.

**Theorem 3.1** *Under Assumption 3.1, the sequence generated by Algorithm 2.1 converges to some solution of (1.1) quadratically.*

*Proof* By the SVD of $J_k$,

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

and

$$F_k + J_kd_k = \lambda_kU_1(\Sigma_1^2+\lambda_kI)^{-1}U_1^TF_k+\lambda_kU_2(\Sigma_2^2+\lambda_kI)^{-1}U_2^TF_k.$$

It follows from (3.4), (3.23), Lemmas 3.6 and 3.7 that

$$\begin{aligned}\|F_k + J_kd_k\| \le&\mu_k\|F_k\|^2\|\Sigma_1^{-2}\|\|U_1U_1^TF_k\| + \|U_2U_2^TF_k\|\\ \le&\frac{4L_2^2C_3}{\bar{\sigma}_r^2}\|\bar{x}_k - x_k\|^2 + 2L_1\|\bar{x}_k - x_k\|^2\\ \le&c_5\|\bar{x}_k - x_k\|^2,\end{aligned}$$

where $c_5 = \frac{4L_2^2C_3}{\bar{\sigma}_r^2} + 2L_1$ is a positive constant. So, by (3.1), (3.3) and Lemma 3.1,

$$\begin{aligned}c\|\bar{x}_{k+1} - x_{k+1}\| \le& \|F_{k+1}\|\\ \le& \|F_k + J_kd_k\| + L_1\|d_k\|^2\\ \le& c_5\|\bar{x}_k - x_k\|^2 + L_1c_2^2\|\bar{x}_k - x_k\|^2\\ \le& c_6\|\bar{x}_k - x_k\|^2,\end{aligned} \tag{3.24}$$

where $c_6 = c_5 + c_2^2L_1$ is a positive constant.
Note that

$$\|\bar{x}_k - x_k\| \le \|\bar{x}_{k+1} - x_{k+1}\| + \|d_k\|. \tag{3.25}$$

By (3.24),

$$\|\bar{x}_k - x_k\| \le 2\|d_k\|$$

holds for all sufficiently large $k$. Combining (3.5), (3.24) and (3.25), we obtain

$$\|d_{k+1}\| \le O(\|d_k\|^2).$$

The proof is completed. □

**Table 1** Results on the first singular test set with $\text{rank}(F'(x^*)) = n - 1$

| Prob | $n$ | $x_0$ | $\lambda_k = \mu_k\|F_k\|$ with (1.7) NF/NJ/NF+NJ*$n$ | $\lambda_k = \mu_k\|F_k\|^2$ with (1.7) NF/NJ/NF+NJ*$n$ | $\lambda_k = \mu_k\|F_k\|^2$ with (1.9) NF/NJ/NF+NJ*$n$ |
|---|---|---|---|---|---|
| 1 | 2 | 1 | 16/16/48 | 16/16/48 | 16/16/48 |
| | | 10 | 19/19/57 | 19/19/57 | 19/19/57 |
| | | 100 | 22/22/66 | 23/23/69 | 23/23/69 |
| 4 | 4 | 1 | 18/18/90 | 18/18/90 | 18/18/90 |
| | | 10 | 20/20/100 | 20/20/100 | 20/20/100 |
| | | 100 | 24/24/120 | 24/24/120 | 24/24/120 |
| 5 | 3 | 1 | 8/8/32 | 8/8/32 | 8/8/32 |
| | | 10 | 8/8/32 | 8/8/32 | 8/8/32 |
| | | 100 | 8/8/32 | 8/8/32 | 8/8/32 |
| 6 | 31 | 1 | 156/124/4000 | 68/35/1153 | 737/368/12145 |
| 8 | 10 | 1 | 9/9/99 | 9/9/99 | 9/9/99 |
| | | 10 | 24/24/264 | 24/24/264 | 24/24/264 |
| 9 | 10 | 1 | 5/5/55 | 5/5/55 | 5/5/55 |
| | | 10 | 6/6/66 | 6/6/66 | 6/6/66 |
| | | 100 | 10/10/110 | 10/10/110 | 10/10/110 |
| 10 | 30 | 1 | 7/7/217 | 7/7/217 | 7/7/217 |
| | | 10 | 9/9/279 | 9/9/279 | 9/9/279 |
| | | 100 | 10/10/310 | 10/10/310 | 10/10/310 |
| 11 | 30 | 1 | 31/13/421 | 45/18/585 | 37/14/457 |
| 12 | 10 | 1 | 15/15/165 | 15/15/165 | 15/15/165 |
| | | 10 | 17/17/187 | 17/17/187 | 17/17/187 |
| | | 100 | 21/21/231 | 21/21/231 | 21/21/231 |
| 13 | 30 | 1 | 11/11/341 | 11/11/341 | 11/11/341 |
| | | 10 | 15/15/465 | 15/15/465 | 15/15/465 |
| | | 100 | 19/19/589 | 19/19/589 | 19/19/589 |
| 14 | 30 | 1 | 14/14/434 | 14/14/434 | 14/14/434 |
| | | 10 | 20/20/620 | 20/20/620 | 20/20/620 |
| | | 100 | 26/26/806 | 26/26/806 | 26/26/806 |

*Remark 3.1* If the Levenberg–Marquardt parameter is chosen as $\lambda_k = \mu_k\|F_k\|^\delta$, where $\mu_k$ is updated by (1.9) and $\delta \in (1, 2]$, the algorithm converges superlinearly to some solution of the nonlinear equations with the order $\delta$. The proof is almost the same as above, except that we have $\|F_k + J_k d_k\| \leq c_5\|\bar{x}_k - x_k\|^\delta$ instead of $\|F_k + J_k d_k\| \leq c_5\|\bar{x}_k - x_k\|^2$ in the proof of Theorem 3.1, which then yields $\|d_{k+1}\| \leq O(\|d_k\|^\delta)$.

## 4 Numerical Results

We test Algorithm 2.1, where the LM parameter is computed by $\lambda_k = \mu_k\|F_k\|^2$ with $\mu_k$ updated by (1.9), on some singular nonlinear equations, and compare it with other two LM algorithms, where $\lambda_k = \mu_k\|F_k\|$ and $\lambda_k = \mu_k\|F_k\|^2$ with $\mu_k$ updated by (1.7), respectively.

**Table 2** Results on the second singular test set with rank$(F'(x^*)) = n - 2$

| Prob | $n$ | $x_0$ | $\lambda_k = \mu_k \|F_k\|$ with (1.7) NF/NJ/NF+NJ*$n$ | $\lambda_k = \mu_k \|F_k\|^2$ with (1.7) NF/NJ/NF+NJ*$n$ | $\lambda_k = \mu_k \|F_k\|^2$ with (1.9) NF/NJ/NF+NJ*$n$ |
|---|---|---|---|---|---|
| 1 | 2 | 1 | 12/12/36 | 12/12/36 | 12/12/36 |
| | | 10 | 14/14/42 | 14/14/42 | 14/14/42 |
| | | 100 | 18/18/54 | 18/18/54 | 18/18/54 |
| 3 | 2 | 1 | 38/26/90 | 211/138/487 | 260/180/620 |
| | | 10 | 30/15/60 | 37/15/67 | 19/12/43 |
| | | 100 | 34/18/70 | 39/16/71 | 22/16/54 |
| 4 | 4 | 1 | 15/15/75 | OF | 15/15/75 |
| | | 10 | 18/18/90 | OF | 18/18/90 |
| | | 100 | 22/22/110 | OF | 22/22/110 |
| 5 | 3 | 1 | 15/15/60 | 15/15/60 | 15/15/60 |
| | | 10 | 16/16/64 | 16/16/64 | 16/16/64 |
| | | 100 | 16/16/64 | 16/16/64 | 16/16/64 |
| 6 | 31 | 1 | 3200/2621/84451 | –/–/– | 1288/665/21903 |
| 8 | 10 | 1 | 9/9/99 | 9/9/99 | 9/9/99 |
| | | 10 | 23/23/253 | 24/24/264 | 24/24/264 |
| 9 | 10 | 1 | 5/5/55 | 5/5/55 | 5/5/55 |
| | | 10 | 8/8/88 | 8/8/88 | 8/8/88 |
| | | 100 | 10/10/110 | 10/10/110 | 10/10/110 |
| 10 | 30 | 1 | 7/7/217 | 7/7/217 | 7/7/217 |
| | | 10 | 10/10/310 | 9/9/279 | 9/9/279 |
| | | 100 | 10/10/310 | 11/11/341 | 22/18/562 |
| 11 | 30 | 1 | 30/13/420 | 36/14/456 | 65/44/1385 |
| 12 | 10 | 1 | 15/15/165 | 15/15/165 | 15/15/165 |
| | | 10 | 17/17/187 | 17/17/187 | 17/17/187 |
| | | 100 | 21/21/589 | 21/21/589 | 21/21/589 |
| 13 | 30 | 1 | 11/11/341 | 11/11/341 | 11/11/341 |
| | | 10 | 15/15/465 | 15/15/465 | 15/15/465 |
| | | 100 | 19/19/589 | –/–/– | 19/19/589 |
| 14 | 30 | 1 | 14/14/434 | 14/14/434 | 14/14/434 |
| | | 10 | 20/20/620 | 19/19/589 | 19/19/589 |
| | | 100 | 26/26/806 | 25/25/775 | 26/26/806 |

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

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

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

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

**Table 3** Results on the first singular test set with rank $n-1$

| Prob | $n$ | $x_0$ | $\lambda_k = \mu_k \|F_k\|$ with (1.7) NF/NJ/NF+NJ*$n$/t(s) | $\lambda_k = \mu_k \|F_k\|^2$ with (1.7) NF/NJ/NF+NJ*$n$/t(s) | $\lambda_k = \mu_k \|F_k\|^2$ with (1.9) NF/NJ/NF+NJ*$n$/t(s) |
|---|---|---|---|---|---|
| 9 | 3000 | 1 | 1/1/3001/0.3 | 1/1/3001/0.3 | 1/1/3001/0.3 |
| | | 10 | 2/2/6002/20.3 | 2/2/6002/19.3 | 2/2/6002/19.5 |
| | | 100 | 8/8/24008/147 | 8/8/24008/131 | 8/8/24008/128 |
| 10 | 3000 | 1 | 18/9/27018/593 | 27/9/27027/714 | 20/9/27020/562 |
| | | 10 | 20/11/33020/645 | 29/11/33029/848 | 20/11/33020/623 |
| | | 100 | 10/10/30010/415 | 10/10/30010/415 | 10/10/30010/426 |
| 13 | 3000 | 1 | 11/11/33011/220 | 11/11/33011/220 | 11/11/33011/218 |
| | | 10 | 15/15/45015/289 | 15/15/45015/302 | 16/16/48016/298 |
| | | 100 | 19/19/57019/388 | 19/19/57019/393 | 19/19/57019/379 |
| 14 | 3000 | 1 | 14/14/42014/386 | 14/14/42014/302 | 14/14/42014/270 |
| | | 10 | 20/20/60020/588 | 20/20/60020/446 | 20/20/60020/412 |
| | | 100 | 26/26/78026/789 | 26/26/78026/584 | 26/26/78026/557 |

**Table 4** Results on the first singular test set with rank $n-2$

| Prob | $n$ | $x_0$ | $\lambda_k = \mu_k \|F_k\|$ with (1.7) NF/NJ/NF+NJ*$n$/t(s) | $\lambda_k = \mu_k \|F_k\|^2$ with (1.7) NF/NJ/NF+NJ*$n$/t(s) | $\lambda_k = \mu_k \|F_k\|^2$ with (1.9) NF/NJ/NF+NJ*$n$/t(s) |
|---|---|---|---|---|---|
| 9 | 3000 | 1 | 1/1/3001/0.4 | 1/1/3001/0.3 | 1/1/3001/0.3 |
| | | 10 | 2/2/6002/21.3 | 2/2/6002/20.3 | 2/2/6002/20.1 |
| | | 100 | 10/10/30010/200 | 9/9/27009/169 | 6/6/18006/103 |
| 10 | 3000 | 1 | 18/9/27018/591 | 29/10/30029/831 | 20/9/27020/620 |
| | | 10 | 20/11/33020/696 | 29/11/33029/1039 | 20/11/33020/604 |
| | | 100 | 25/15/45025/903 | 34/15/45034/1270 | 10/10/30010/364 |
| 13 | 3000 | 1 | 11/11/33011/203 | 11/11/33011/229 | 11/11/33011/202 |
| | | 10 | 15/15/45015/303 | 15/15/45015/299 | 16/16/48016/306 |
| | | 100 | 19/19/57019/402 | 19/19/57019/402 | 19/19/57019/374 |
| 14 | 3000 | 1 | 14/14/42014/398 | 14/14/42014/367 | 14/14/42014/308 |
| | | 10 | 20/20/60020/441 | 20/20/60020/444 | 20/20/60020/454 |
| | | 100 | 26/26/78026/579 | 26/26/78026/599 | 26/26/78026/597 |

has rank $n-k$. A disadvantage of these problems is that $\hat{F}(x)$ may have roots that are not roots of $F(x)$. We create two sets of singular problems, with $\hat{J}(x^*)$ having rank $n-1$ and $n-2$, by using

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

and

$$A \in R^{n\times 2}, \qquad A^T = \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & -1 & 1 & -1 & \cdots & \pm 1 \end{pmatrix},$$

respectively. Meanwhile, we make a slight alteration on the variable dimension problem, which has $n+2$ equations in $n$ unknowns; we eliminate the $(n-1)$-th and $n$-th equations. (The first $n$ equations in the standard problem are linear.)

We set $p_0 = 0.0001, p_1 = 0.25, p_2 = 0.75, c_0 = 4, c_1 = 0.25, \mu_0 = 10^{-8}, m = 10^{-8}, \varepsilon = 10^{-6}$ for all the tests. The stopping criterion is $\|J_k^T F_k\| \le \varepsilon$ or when the number of iterations exceeds $100(n+1)$. The results for the first set problems of rank $n-1$ with small scale are listed in Table 1, and the second set of rank $n-2$ in Table 2. We also test the algorithms on some large scale problems. The results are given in Tables 3 and 4.

The third column of the table indicates that the starting point is $x_0$, $10x_0$, and $100x_0$, where $x_0$ is suggested by Moré, Garbow and Hillstrom in [8]; "NF" and "NJ" represent the numbers of function calculations and Jacobian calculations, respectively. If the algorithm fails to find the solution in $100(n+1)$ iterations, we denote it by the sign "−", and if the algorithm has underflows or overflows, we denote it by OF. Note that, for general nonlinear equations, the calculations of the Jacobian are usually $n$ times of the function calculations. So, for small scale problems, we also present the values "NF+$n$*N" for comparisons of the total calculations. However, if the Jacobian is sparse, this kind of value does not mean much. For the large scale problem, the computing time is also given.

From Tables 1 and 2, we can see that Algorithm 2.1 works almost the same as other two LM algorithms for small scale problems. From Tables 3 and 4, we can see that Algorithm 2.1 outperforms the other two algorithms for most large scale problems.

## 5 Conclusion and Discussion

In traditional LM algorithms for nonlinear equations, both the iterate and the LM parameter are updated according to the ratio of the actual reduction to the predicted reduction of the merit function (cf. [1,2]). In this paper, we proposed a new LM algorithm for nonlinear equations, where the LM parameter is taken as $\lambda_k = \mu_k \|F_k\|^2$ with $\mu_k$ being updated by (1.9). Though the iterate is still updated according to the ratio of the actual reduction to the predicted reduction, the update of $\mu_k$ is no longer based on it. When the iteration is unsuccessful, $\mu_k$ is increased; otherwise it is updated based on the value of the gradient norm of the merit function as in (1.9). We proved that all limit points of the sequence generated by the algorithm are stationary points of the merit function under standard conditions. Since the updating rule of $\mu_k$ changes, the analysis of the convergence rate in this paper is quite different from those in [1,3]. We developed new techniques to prove the quadratic convergence of the algorithm under the local error bound condition.

We also discussed the LM parameter as $\lambda_k = \mu_k \|F_k\|^\delta$, where $\mu_k$ is updated by (1.9) and $\delta \in [1, 2)$. We found that the algorithm converges with the order $\delta$, by using the similar analysis in this paper. We conjecture that the convergence rate is quadratic for any $\delta \in [1, 2)$. This will be our future study.

## References

1. Fan, J.Y.: A modified Levenberg–Marquardt algorithm for singular system of nonlinear equations. J. Comput. Math. **21**, 625–636 (2003)
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., Pan, J.Y., Song, H.Y.: A retrospective trust region algorithm with trust region converging to zero. J. Comput. Math. **34**, 421–436 (2016)
4. Fan, J.Y., Yuan, Y.X.: On the quadratic convergence of the Levenberg–Marquardt method without non-singularity assumption. Computing **74**, 23–39 (2005)

5. Levenberg, K.: A method for the solution of certain nonlinear problems in least squares. Quardt. Appl. Math. **2**, 164–166 (1944)
6. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear inequalities. SIAM J. Appl. Math. **11**, 431–441 (1963)
7. Moré, J.J.: The Levenberg–Marquardt algorithm: implementation and theory. Numer. Anal. **630**, 105–116 (1978)
8. Moré, J.J., Garbow, B.S., Hillstrom, K.H.: Testing unconstrained optimization software. ACM Trans. Math. Softw. **7**, 17–41 (1981)
9. Osborne, M.R.: Nonlinear least squares-the Levenberg–Marquardt algorithm revisited. J. Aust. Math. Soc. **19**, 343–357 (1976)
10. Powell, M.J.D.: Convergence properties of a class of minimization algorithms. In: Mangasarian, O.L., Meyer, R.R., Robinson, S.M. (eds.) Nonlinear Programming 2, pp. 1–27. Academic Press, New York (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. Academic Press, San Diego (1990)
13. Wright, S.J., Holt, J.N., Holt, J.N.: An inexact Levenberg–Marquardt method for large sparse nonlinear least squares. J. Aust. Math. Soc. **26**, 387–403 (1985)
14. Yamashita, N., Fukushima, M.: On the rate of convergence of the Levenberg–Marquardt mehod. Computing (Supplement) **15**, 237–249 (2001)
15. Yuan, Y.X.: Trust region algorithms for nonlinear equations. Information **1**, 7–20 (1998)
16. Yuan, Y.X.: Recent advances in numerical methods for nonlinear equations and nonlinear least sqaures. Numer. Algebra Control Optim. **1**, 15–34 (2011)
17. Yuan, Y.X.: Recent advances in trust region algorithms. Math. Program. Ser. B **151**, 249–281 (2015)
18. Zhao, R.X., Fan, J.Y.: Global complexity bound of the Levenberg–Marquardt method. Optim. Methods Softw. **31**, 805–814 (2016)