Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 21, No. 1, pp. 63–87, 2025

A NONMONOTONE LEVENBERG-MARQUARDT METHOD WITH CORRECTION FOR SOLVING SYSTEMS OF NONLINEAR EQUATIONS*

Baohua Huang and Changfeng Ma

Abstract: It is well known that Levenberg-Marquardt (LM) method is widely used for solving nonlinear equations. In this paper, we give an extension of LM method and propose a nonmonotone LM method with correction which produces the LM parameter according to the new nonmonotone strategy of Grippo, Lampariello and Lucidi. Moreover, not only an LM step but also a correction step are computed at every iteration in our proposed nonmonotone LM method with correction. The cubic convergence of the proposed method is proved under the local error bound condition which is weaker than nonsingularity. Some numerical results confirm the feasibility and effectiveness of the proposed algorithm.

Key words: *nonlinear equations, Levenberg-Marquardt method, nonmonotone technique, correction step, local error bound condition*

Mathematics Subject Classification: *65H10, 65K05, 90C30*

1 Introduction

In this paper, we consider the following system of nonlinear equations

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

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable mapping. Throughout this paper, we assume that the solution set X^* of (1.1) is nonempty since (1.1) may have no solutions for the nonlinearity of $F(x)$, and in all cases $\|\cdot\|$ stands for the 2-norm. A lot of efficient iterative algorithms have been proposed for the nonlinear equations (1.1), including Newton method, quasi-Newton method, Gauss-Newton method, trust region method, tensor method, Levenberg-Marquardt method, and so on. One can refer to [1, 2, 4–7, 9, 11–15, 18, 20–22, 24, 25, 33–37].

The Levenberg-Marquardt (LM) method is one of the most popular methods for the nonlinear equations (1.1), which computes the search direction by

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

where $F_k = F(x_k)$, $J_k = F'(x_k)$ is the Jacobian of F at x_k, and λ_k is a nonnegative regularized parameter that is updated at each iteration. The LM step (1.2) is actually

*The work is supported by National Natural Science Foundation of China (12001211, 12071159, 12171168) and Natural Science Foundation of Fujian Province, China (2022J01194).

 DOI: https://doi.org/10.61208/pjo-2023-026

a modification of Gauss-Newton step $d_k^{\mathrm{GN}} = -(J_k^T J_k)^{-1} J_k^T F_k$. The LM parameter λ_k is introduced to overcome the difficulty when $J_k^T J_k$ is singular or nearly singular. It is easy to see that the LM step d_k reduces to the Newton step $d_k^{\mathrm{N}} = -J_k^{-1} F_k$ when J_k is nonsingular and $\lambda_k = 0$.

There are many ways to select LM parameter λ_k in (1.2). For instance, Yamashita and Fukushima [34] adopted the LM parameter $\lambda_k = \|F_k\|^2$ and proved that the LM method has quadratic convergence under the local error bound condition. Another LM parameter is chosen as $\lambda_k = \theta\|F_k\| + (1-\theta)\|J_k^T F_k\|$ with θ being a constant in $[0,1]$, which is the convex combination of $\|F_k\|$ and $\|J_k^T F_k\|$ [27]. A more general choice of LM parameter is proposed by Fan and Pan [16], which has the following form

$$\lambda_k = \mu_k \rho(x_k), \tag{1.3}$$

where μ_k is updated at each iteration by the trust region technique and

$$\rho(x_k) = \begin{cases} \tilde{\rho}(x_k), & \text{if } \tilde{\rho}(x_k) \le 1, \\ 1, & \text{otherwise,} \end{cases} \quad \text{with } \tilde{\rho}(x_k) = O(\|F_k\|^\delta). \tag{1.4}$$

Another popular approach employed in the selection of the LM parameter is self-adaptive technique, e.g., Fan and Pan [17] chose the following self-adaptive LM parameter

$$\lambda_k = \mu_k \|F_k\|^\delta, \text{ with } \mu_{k+1} = \mu_k q(r_k), \tag{1.5}$$

where $q(r)$ is a continuous nonnegetive function of r and $\delta \in (0,2]$. Here μ_k is updated at a variable rate according to the ratio r_k, rather than by simply enlarging or reducing the original one at a constant rate. It is well known that LM method is closely related to the trust region method and LM parameter can be updated by using trust region techniques. Recently, Esmaeili and Kimiaei [11, 12] introduced a new adaptive trust region radius Δ_k by using the following formula

$$\Delta_k = c^{p_k} \max\{\Delta_{k-1}, \Lambda_k\},$$

where p_k is a nonnegative integer, and Λ_k is generated by the nonmonotone technique of Grippo, Lampariello and Lucidiv [21, 22] and will be introduced in Section 2. The interesting question is whether we could employ the nonmonotone technique to produce the LM parameter. This is one of our motivations.

We all know that the LM method achieves quadratic convergence when the Jacobian is Lipschitz continuous and nonsingular at the solution. Fan and Yuan also proved in [18] that the LM method preserves the quadratic convergence when $\lambda_k = \|F_k\|^\delta$ for any $\delta \in [1,2]$. Obviously the cost of calculations will be expensive when the dimension of the nonlinear equations (1.1) is large. To save calculations, Fan [35] proposed the modified LM method (MLM) by computing an approximate LM step

$$d_k^{\mathrm{MLM}} = -(J_k^T J_k + \lambda_k I)^{-1} J_k^T F(y_k), \tag{1.6}$$

where $y_k = x_k + d_k$, and set the trial step be $s_k^{\mathrm{MLM}} = d_k + d_k^{\mathrm{MLM}}$. Later, Fan [14] introduced an accelerated MLM method (AMLM) by using a line search strategy to generate a modified LM step and showed that the convergence rate of the algorithm is $\min\{1+2\delta, 3\}$, which results the cubic convergence for $\delta \ge 1$. Following the idea of Shamanskii [31], Huang and Ma [23] proposed a Shamanskii-like self-adaptive LM method (SALM) for nonlinear equations, to avoid more Jacobian calculations and save the linear algebra work as well. The

Shamanskii-like self-adaptive LM method in [23] employs the self-adaptive LM parameter with the form of

$$\lambda_k = \mu_k \rho(x_k)$$

where $\rho(x_k)$ is defined as in (1.4) and μ_k is updated by the following formula

$$\mu_{k+1} = \max\left\{\mu, \mu_k \max\left\{\frac{1}{4}, 1 - 2(2r_k - 1)^3\right\}\right\} \tag{1.7}$$

with r_k being the ratio of the actual reduction to the predicted reduction at the k-th iteration. The Shamanskii-like self-adaptive LM method achieves the global convergence and has the $(m+1)$-order convergence rate under the error bound condition.

The MLM, AMLM and SALM methods modify the LM method by computing a few approximate LM steps for nonlinear equations, to avoid more Jacobian calculations. Different from the above methods, Fan and Zeng [19] gave another classical modification of LM method by computing a correction LM step, called LM method with correction. At each iteration, the algorithm in [19] firstly obtained d_k by solving the following linear equation

$$(J_k^T J_k + \lambda_k I) d_k = -J_k^T F_k, \quad \lambda_k = \mu_k \|F_k\|^\delta, \tag{1.8}$$

where $\delta \in (0, 2]$ and $\mu_k > 0$ is updated by the trust region technique. Then they solve the linear equation

$$(J_k^T J_k + \lambda_k I)\widetilde{d}_k = \lambda_k d_k \tag{1.9}$$

to get the correction step $\widetilde{d}_k = (J_k^T J_k + \lambda_k I)^{-1} \lambda_k d_k$ and set $s_k = d_k + \widetilde{d}_k$ as the search direction. Under the local error bound condition, they proved that the convergence rate of the correction LM method is $\min\{2, 1 + 2\delta\}$. Now a natural question arises: Is it possible to modify the LM method by computing both the approximate LM step and the correction LM step for the nonlinear equations?

By these motivations, in this paper, we propose a new LM algorithm that produces the LM parameter at each iteration by the nonmonotone technique of Grippo, Lampariello and Lucidiv [21, 22]. The new nonmonotone LM parameter is a modification of the adaptive trust region radius with the nonmonotone technique in [11, 12], and is different from the self-adaptive LM parameter in [23]. Then we integrate the approximate LM step and the correction LM step into the new LM algorithm for obtaining the better numerical performance. In contrast to the LM method with correction of Fan and Zeng [19], our proposed LM algorithm employs not only the correction LM step, but also the approximate LM step. It is shown by Tables 1, 2 and 3 in Section 5 that the proposed new algorithm retains the quick convergence of LM method, while significantly decreasing the computational costs of the method due to improving the LM parameter and integrating the approximate LM step and correction LM step.

The main contributions of this paper are given below.

- We propose a new LM algorithm by producing the LM parameter using the nonmonotone technique. The trial steps in our proposed LM algorithm consist of the classical LM step d_k in (1.2), the approximate LM step d_k^{MLM} in (1.6) and an additional correction LM step $\tilde{d}_k = (J_k^T J_k + \lambda_k I)^{-1} \lambda_k d_k^{\mathrm{MLM}}$.
- We investigate the global convergence of the proposed LM algorithm and establish its cubic convergence properties under the local error bound condition.

The remainder of this paper is organized as follows. In Section 2, we propose a nonmonotone LM method with correction in which the LM parameter is generated by using the nonmonotone technique and the trial step is produced by integrating the approximate LM step and the correction LM step into the standard LM step. In Section 3, we give the global convergence of the proposed algorithm under some suitable assumptions. In Section 4, we obtain the convergence order of the new algorithm under the local error bound condition. The numerical experiments of the proposed algorithm are shown and analyzed in Section 5. The paper ends up with some conclusions in Section 6.

2 Nonmonotone Levenberg-Marquardt Method with Correction

It is known that LM method can achieve the global convergence by integrating the trust region technique into the update of LM parameter. In [11, 12], Esmaeili and Kimiaei introduced an adaptive trust region radius based on the nonmonotone technique of Grippo, Lampariello and Lucidiv [21, 22]. Motivated by their work, we hope to produce the new LM parameter by using the nonmonotone technique. More concretely, we modify the LM parameter λ_k by constructing the following quantity

$$\Lambda_k = \begin{cases} \|F_k\|^\delta, & \text{if } k = 0, \\ \dfrac{\sum_{i=0}^{m(k)-1} \eta^{m(k)-i} \mathcal{F}_k(i) + \|F_k\|^\delta}{\sum_{i=0}^{m(k)-1} \eta^{m(k)-i} + 1}, & \text{if } k > 0, \end{cases} \tag{2.1}$$

where $m(0) = 0$, $0 \le m(k) \le \min\{m(k-1)+1, N\}$, $\eta \in [\eta_{\min}, \eta_{\max}]$, $\eta_{\min} \in [0, 1)$, $\eta_{\max} \in [\eta_{\min}, 1]$, $\delta \in [1, 2]$ and

$$\mathcal{F}_k(i) = \begin{cases} \|F_k\|^\delta, & \text{if } k < N, \\ \|F_{k-N+i+1}\|^\delta, & \text{if } k \ge N, \end{cases} \tag{2.2}$$

where

$$i \in \begin{cases} [0, k], & \text{if } k < N, \\ [0, \text{N-1}], & \text{if } k \ge N. \end{cases} \tag{2.3}$$

For the convenience, we also denote

$$\mathcal{F}_k = \{\|F_{k-j}\|^\delta\}_{0 \le j \le m(k)}, \ k \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}. \tag{2.4}$$

Now we produce the following nonmonotone LM parameter

$$\lambda_k = \mu_k \Lambda_k, \tag{2.5}$$

where μ_k is updated by simply enlarging or reducing the original one at a constant rate.

On the basis of the above discussion, the nonmonotone LM method with correction can be outlined as follows.

Algorithm 2.1. (Nonmonotone LM method with correction)

Step 1 Choose the initial point $x_0 \in \mathbb{R}^n$ and several constants $\varepsilon \ge 0$, $\mu_0 > \mu > 0$, $0 < p_0 \le p_1 \le p_2 < 1$, $N > 0$, $\eta \in [\eta_{\min}, \eta_{\max}]$, $\delta \in [1, 2]$.

Step 2 Let $\Lambda_0 = \|F_0\|^\delta$ and $\lambda_0 = \mu_0 \Lambda_0$. Set $m(0) = 0$ and $k = 0$.

Step 3 Compute $F_k = F(x_k)$ and $J_k = J(x_k)$. If $\|J_k^T F_k\| \le \varepsilon$, then stop. Otherwise, compute d_k by solving

$$(J_k^T J_k + \lambda_k I)d = -J_k^T F_k \text{ with } \lambda_k = \mu_k \Lambda_k. \tag{2.6}$$

Let $y_k = x_k + d_k$. Compute $\hat{d}_k$ by solving

$$(J_k^T J_k + \lambda_k I)d = -J_k^T F(y_k). \tag{2.7}$$

Then solve

$$(J_k^T J_k + \lambda_k I)d = -J_k^T F(y_k) + \lambda_k \hat{d}_k \tag{2.8}$$

to obtain $\tilde{d}_k$. Set

$$s_k = d_k + \tilde{d}_k. \tag{2.9}$$

Step 4 Compute $r_k = Ared_k/Pred_k$. Set

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{if } r_k \geq p_0, \\ x_k, & \text{otherwise.} \end{cases} \tag{2.10}$$

Step 5 Choose $m(k+1) \in \big[0, \min\{m(k)+1, N\}\big]$. Compute Λ_{k+1} by using (2.1).

Step 6 Choose μ_{k+1} by the following formula

$$\mu_{k+1} = \begin{cases} 4\mu_k, & \text{if } r_k < p_1, \\ \mu_k, & \text{if } r_k \in [p_1, p_2], \\ \max\{\mu, \dfrac{\mu_k}{4}\}, & \text{if } r_k > p_2. \end{cases} \tag{2.11}$$

Set $k = k+1$, go to Step 3.

Remark 2.2. The given positive constant μ in Algorithm 2.1 is the lower bound of the LM parameter which prevents the step from being too large in the case that the sequence is near the solution.

So far, the quantity r_k in Step 4 of Algorithm 2.1 is still unclear. And for that, we are going to give the definitions of $Ared_k$ and $Pred_k$.

First we take

$$\Psi(x) = \|F(x)\|^2 \tag{2.12}$$

as the merit function for (1.1). The actual reduction of $\Psi(x)$ at the kth iteration is defined by

$$Ared_k = \|F_k\|^2 - \|F(x_k + s_k)\|^2, \tag{2.13}$$

where s_k is the trial step defined as in (2.9).

Since $\|F_k\|^2 - \|F_k + J_k s_k\|^2$ can not be proved to be nonnegative, we can not define it as the predicted reduction as usual. Hence a modified predicted reduction is need to be given.

Note that d_k is not only the minimizer of the convex minimization problem

$$\min_{d\in\mathbb{R}^n} \|F_k + J_k d\|^2 + \lambda_k \|d\|^2 \triangleq \varphi_{k,1}(d), \tag{2.14}$$

but also a solution of the following trust region problem

$$\begin{aligned} &\min_{d\in\mathbb{R}^n} \|F_k + J_k d\|^2 \\ &\text{s.t. } \|d\| \leq \triangle_{k,1}, \end{aligned} \tag{2.15}$$

where

$$\triangle_{k,1} = \| - (J_k^T J_k + \lambda_k I)^{-1} J_k^T F_k\| = \|d_k\|.$$

According to the result developed by Powell [29], it follows that

$$\|F_k\|^2 - \|F_k + J_k d_k\|^2 \geq \|J_k^T F_k\| \min\Big\{\|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\Big\}. \tag{2.16}$$

It is also easy to see that $\hat{d}_k$ is not only the minimizer of the convex minimization problem

$$\min_{d\in\mathbb{R}^n} \|F(y_k) + J_k d\|^2 + \lambda_k \|d\|^2 \triangleq \varphi_{k,2}(d), \tag{2.17}$$

but also a solution of the following trust region problem

$$\begin{array}{l} \min\limits_{d\in\mathbb{R}^n} \|F(y_k) + J_k d\|^2 \triangleq \psi_k(d) \\ s.t.\ \|d\| \leq \triangle_{k,2}, \end{array} \tag{2.18}$$

where

$$\triangle_{k,2} = \| - (J_k^T J_k + \lambda_k I)^{-1} J_k^T F(y_k)\| = \|\hat{d}_k\|. \tag{2.19}$$

So we also have

$$\psi_k(0) - \psi_k(\hat{d}_k) = \|F(y_k)\|^2 - \|F(y_k) + J_k \hat{d}_k\|^2 \geq \|J_k^T F(y_k)\| \min\Big\{\|\hat{d}_k\|, \frac{\|J_k^T F(y_k)\|}{\|J_k^T J_k\|}\Big\}. \tag{2.20}$$

In order to define the predicted reduction, we call

$$d_k^c = (J_k^T J_k + \lambda_k I)^{-1} \lambda_k \hat{d}_k \tag{2.21}$$

the correction step. Then it follows from (2.7) and (2.8) that

$$\tilde{d}_k = \hat{d}_k + d_k^c. \tag{2.22}$$

By using the relations (2.7), (2.8), (2.18), (2.21) and (2.22), we obtain

$$\begin{aligned} \psi_k(\hat{d}_k) - \psi_k(\tilde{d}_k) &= \|F(y_k) + J_k \hat{d}_k\|^2 - \|F(y_k) + J_k \tilde{d}_k\|^2 \\ &= 2\hat{d}_k^T J_k^T F(y_k) + \hat{d}_k^T J_k^T J_k \hat{d}_k - 2\tilde{d}_k^T J_k^T F(y_k) - \tilde{d}_k^T J_k^T J_k \tilde{d}_k \\ &= -2(d_k^c)^T J_k^T F(y_k) - (d_k^c)^T J_k^T J_k (d_k^c) - 2(d_k^c)^T J_k^T J_k \hat{d}_k \\ &= 2(d_k^c)^T (J_k^T J_k + \lambda_k I)\hat{d}_k - (d_k^c)^T J_k^T J_k (d_k^c) - 2(d_k^c)^T J_k^T J_k \hat{d}_k \\ &= 2\lambda_k (d_k^c)^T \hat{d}_k - (d_k^c)^T J_k^T J_k (d_k^c) \\ &= 2(d_k^c)^T (J_k^T J_k + \lambda_k I)(d_k^c) - (d_k^c)^T J_k^T J_k (d_k^c) \\ &= 2\lambda_k (d_k^c)^T (d_k^c) + (d_k^c)^T J_k^T J_k (d_k^c) \\ &\geq 0. \end{aligned} \tag{2.23}$$

This together with (2.20) yields

$$\begin{aligned} \|F(y_k)\|^2 - \|F(y_k) + J_k \tilde{d}_k\|^2 &= \psi_k(0) - \psi_k(\tilde{d}_k) \\ &= [\psi_k(0) - \psi_k(\hat{d}_k)] + [\psi_k(\hat{d}_k) - \psi_k(\tilde{d}_k)] \\ &\geq \psi_k(0) - \psi_k(\hat{d}_k) \\ &\geq \|J_k^T F(y_k)\| \min\Big\{\|\hat{d}_k\|, \frac{\|J_k^T F(y_k)\|}{\|J_k^T J_k\|}\Big\}. \end{aligned} \tag{2.24}$$

Hence the new predicted reduction can de specified by

$$Pred_k = \|F_k\|^2 - \|F_k + J_k d_k\|^2 + \|F(y_k)\|^2 - \|F(y_k) + J_k \tilde{d}_k\|^2. \tag{2.25}$$

By (2.16), (2.24) and (2.25), it follows that

$$\begin{aligned} Pred_k &= \|F_k\|^2 - \|F_k + J_k d_k\|^2 + \|F(y_k)\|^2 - \|F(y_k) + J_k \tilde{d}_k\|^2 \\ &\geq \|J_k^T F_k\| \min\left\{\|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\} > 0. \end{aligned} \tag{2.26}$$

According to the relations (2.13) and (2.26), it is easy to see the following fact.

Remark 2.3. By Step 4 of Algorithm 2.1, it follows that

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

This, together with (2.13) and (2.26), yields

$$\|F_k\|^2 - \|F_{k+1}\|^2 = \begin{cases} Ared_k = r_k \times Pred_k \geq p_0 \times Pred_k > 0, & \text{if } r_k \geq p_0, \\ 0, & \text{otherwise.} \end{cases}$$

Hence $\|F_{k+1}\|^2 \leq \|F_k\|^2$. This implies that the sequence $\{\|F_k\|\}$ generated by Algorithm 2.1 is a nonincreasing sequence.

$\boxed{3}$ Global Convergence of Algorithm 2.1

In this section, we will give the global convergence of Algorithm 2.1. We need the following assumptions.

Assumption 3.1. Let the level set $L(x_0) = \{x \in \mathbb{R}^n : \|F(x)\| \leq \|F(x_0)\|\}$ be bounded for any given $x_0 \in \mathbb{R}^n$.

Assumption 3.2. Both $F(x)$ and its Jacobian $J(x)$ are Lipschitz continuous, i.e., there exist positive constants L_1 and L_2 such that

$$\|J(x) - J(y)\| \leq L_1\|x - y\|, \ \forall x, y \in \mathbb{R}^n \tag{3.1}$$

and

$$\|F(x) - F(y)\| \leq L_2\|x - y\|, \ \forall x, y \in \mathbb{R}^n. \tag{3.2}$$

From (3.1) and (3.2), it is easy to check that

$$\|F(y) - F(x) - J(x)(y - x)\| \leq L_1\|y - x\|^2, \ \forall x, y \in \mathbb{R}^n \tag{3.3}$$

and

$$\|J(x)\| \leq L_2, \ \forall x \in \mathbb{R}^n. \tag{3.4}$$

In order to obtain the global convergence of Algorithm 2.1, we need to prove the following lemma.

Lemma 3.1. *Let* Assumptions 3.1 *and* 3.2 *hold and the sequence* $\{x_k\}$ *be generated by* Algorithm 2.1. *Denote*

$$NF_{l(k)} = \max\{\mathcal{F}_k\} = \max_{0\leq j\leq N}\{\|F_{k-j}\|^{\delta}\},\ k\in\mathbb{N}_0.$$

Then the following statements hold:
(1) $\{\Lambda_k\}_{k\geq 0}$ *is a decreasing sequence.*
(2) $x_k \in L(x_0)$ *for all* $k\in\mathbb{N}_0$.
(3) $\{NF_{l(k)}\}_{k\geq 0}$ *is decreasing and hence converges.*
(4) $\lim_{k\to\infty}\Lambda_k = \lim_{k\to\infty}\|F_k\|^{\delta}$.

Proof. (1) From Remark 2.3, it follows that $\|F_{k+1}\| \leq \|F_k\|$. This, together with the relations (2.1) and (2.2), yields

$$\Lambda_k = \frac{\sum_{i=0}^{m(k)-1}\eta^{m(k)-i}\mathcal{F}_k(i) + \|F_k\|^{\delta}}{\sum_{i=0}^{m(k)-1}\eta^{m(k)-i}+1} \geq \frac{\sum_{i=0}^{m(k)-1}\eta^{m(k)-i}\|F_k\|^{\delta} + \|F_k\|^{\delta}}{\sum_{i=0}^{m(k)-1}\eta^{m(k)-i}+1} = \|F_k\|^{\delta}. \quad (3.5)$$

Using the definition of $\mathcal{F}_k$ gives

$$\mathcal{F}_{k+1}(i) \leq \mathcal{F}_k(i),\ \forall i = 1,2,\cdots,N.$$

Combining this and (3.5) leads to

$$\begin{aligned}
\Lambda_{k+1} &= \frac{\sum_{i=0}^{m(k)}\eta^{m(k)-i+1}\mathcal{F}_{k+1}(i) + \|F_{k+1}\|^{\delta}}{\sum_{i=0}^{m(k)}\eta^{m(k)-i+1}+1}\\
&\leq \frac{\sum_{i=0}^{m(k)-1}\eta^{m(k)-i+1}\mathcal{F}_k(i) + \eta\mathcal{F}_k(m(k)) + \|F_{k+1}\|^{\delta}}{\sum_{i=0}^{m(k)}\eta^{m(k)-i+1}+1}\\
&\leq \frac{\eta\Lambda_k\big[\sum_{i=0}^{m(k)-1}\eta^{m(k)-i}+1\big] + \|F_{k+1}\|^{\delta}}{\sum_{i=0}^{m(k)}\eta^{m(k)-i+1}+1}\\
&\leq \frac{\Lambda_k\sum_{i=0}^{m(k)}\eta^{m(k)-i+1} + \|F_k\|^{\delta}}{\sum_{i=0}^{m(k)}\eta^{m(k)-i+1}+1} \leq \Lambda_k.
\end{aligned}$$

This implies that the sequence $\{\Lambda_k\}_{k\geq 0}$ is decreasing.

(2) We prove this result by induction. For $k=0$, the result is trivial. Suppose that $x_i\in L(x_0)$ for $i=1,2,\cdots,k$, i.e.,

$$\|F_i\|^{\delta} \leq \|F_0\|^{\delta},\ \forall i=1,2,\cdots,k.$$

According to Step 4 of Algorithm 2.1 and the relations (2.26) and (3.5), it follows that

$$\begin{aligned}
\Lambda_k^{2/\delta} - \|F_{k+1}\|^2 &\geq \|F_k\|^2 - \|F_{k+1}\|^2\\
&= Ared_k \geq p_0 Pred_k\\
&\geq p_0\|J_k^T F_k\|\min\left\{\|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\} > 0.
\end{aligned}$$

This implies that

$$\|F_{k+1}\|^{\delta} \leq \Lambda_k. \quad (3.6)$$

From the first claim of this lemma, we immediately have

$$\|F_{k+1}\|^\delta \le \Lambda_k \le \Lambda_{k-1} \le \cdots \le \Lambda_0 = \|F_0\|^\delta.$$

Thus $x_{k+1} \in L(x_0)$ and consequently, the sequence $\{x_k\}_{k\ge 0}$ is contained in $L(x_0)$ by the induction principle.

(3) By the definition of $NF_{l(k)}$, we have

$$\Lambda_k = \frac{\sum_{i=0}^{m(k)-1} \eta^{m(k)-i}\mathcal{F}_k(i) + \|F_k\|^\delta}{\sum_{i=0}^{m(k-1)} \eta^{m(k)-i} + 1} \le \frac{\sum_{i=0}^{m(k)-1} \eta^{m(k)-i} NF_{l(k)} + NF_{l(k)}}{\sum_{i=0}^{m(k-1)} \eta^{m(k)-i} + 1} = NF_{l(k)}. \quad (3.7)$$

Combining (3.6) with (3.7) yields $\|F_{k+1}\|^\delta \le \Lambda_k \le NF_{l(k)} \le \|F_0\|^\delta$ and

$$\begin{aligned} NF_{l(k+1)} &= \max_{0\le j\le N}\{\|F_{k+1-j}\|^\delta\} = \max\{\max_{0\le j\le N-1}\{\|F_{k-j}\|^\delta\}, \|F_{k+1}\|^\delta\} \\ &\le \max\{\max_{0\le j\le N}\{\|F_{k-j}\|^\delta\}, \|F_{k+1}\|^\delta\} = \max\{NF_{l(k)}, \Lambda_k\} \le NF_{l(k)}. \end{aligned}$$

This implies that $\{NF_{l(k)}\}_{k\ge 0}$ is decreasing and bounded, and hence converges.

(4) Taking the limit on both sides of (3.6) and (3.7) yields

$$\lim_{k\to\infty} \|F_k\|^\delta \le \lim_{k\to\infty} \Lambda_k \quad \text{and} \quad \lim_{k\to\infty} \Lambda_k \le \lim_{k\to\infty} NF_{l(k)}. \quad (3.8)$$

According to Lemma 3.2 [1], it follows that $\lim_{k\to\infty} NF_{l(k)} = \lim_{k\to\infty} \|F_k\|^\delta$. This, together with the relation (3.8), yields the desired result. □

Now we discuss the global convergence of Algorithm 2.1.

Theorem 3.2. *Let* Assumptions 3.1 and 3.2 *hold. Then the sequence* $\{x_k\}$ *generated by* Algorithm 2.1 *will terminate in finite iterations or satisfy*

$$\lim_{k\to\infty} \|J_k^T F_k\| = 0. \quad (3.9)$$

Proof. We prove this result by contradiction. If the result is not true, then there exist a positive constant $\varepsilon > 0$ and infinite many k such that

$$\|J_k^T F_k\| \ge \varepsilon. \quad (3.10)$$

This together with (3.4) implies that

$$\|F_k\| \ge L_2^{-1}\varepsilon. \quad (3.11)$$

Let

$$T_1 = \big\{k : \|J_k^T F_k\| \ge \varepsilon\big\} \quad \text{and} \quad T_2 = \Big\{k : \|J_k^T F_k\| \ge \frac{\varepsilon}{2} \text{ and } x_{k+1} \ne x_k\Big\}.$$

Obviously T_1 is infinite, here we consider T_2 in two cases.

Case 1: T_2 is infinite. According to Remark 2.3 and the relations (2.26) and (3.4), it follows that

$$\begin{aligned} \|F_1\|^2 &\ge \sum_k \|F_k\|^2 - \|F_{k+1}\|^2 \ge \sum_{k\in T_2} \|F_k\|^2 - \|F_{k+1}\|^2 \ge \sum_{k\in T_2} p_0 Pred_k \\ &\ge \sum_{k\in T_2} p_0 \|J_k^T F_k\| \min\Big\{\|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\Big\} \ge \sum_{k\in T_2} p_0 \frac{\varepsilon}{2} \min\Big\{\|d_k\|, \frac{\varepsilon}{2L_2^2}\Big\}. \quad (3.12) \end{aligned}$$

This means that

$$\sum_{k\in T_2} \|d_k\| < +\infty, \tag{3.13}$$

and then

$$\|d_k\| \to 0,\ \ k \in T_2. \tag{3.14}$$

By the definition of d_k, we obtain

$$\lambda_k \to +\infty,\ \ k \in T_2. \tag{3.15}$$

Using the relations (3.3) and (3.4) yields

$$\begin{aligned}
\|\hat{d}_k\| &= \| -(J_k^T J_k + \lambda_k I)^{-1} J_k^T F(y_k)\| \\
&= \big\| -(J_k^T J_k + \lambda_k I)^{-1} J_k^T \big[\big(F(y_k) - F_k - J_k(y_k - x_k)\big) + F_k + J_k(y_k - x_k)\big]\big\| \\
&\le L_1 \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T\| \|d_k\|^2 + \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T F_k\| \\
&\quad + \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T J_k d_k\| \\
&\le \frac{L_1 L_2}{\lambda_k} \|d_k\|^2 + \|d_k\| + \|d_k\| \\
&\le c_1 \|d_k\|,
\end{aligned} \tag{3.16}$$

where c_1 is a positive number.

On the other hand, from the definition of the correction step d_k^c, we have

$$\|d_k^c\| \le \lambda_k \|(J_k^T J_k + \lambda_k I)^{-1}\| \|\hat{d}_k\| \le \|\hat{d}_k\|, \tag{3.17}$$

and then

$$\|\tilde{d}_k\| = \|\hat{d}_k + d_k^c\| \le 2\|\hat{d}_k\|. \tag{3.18}$$

Thus

$$\|s_k\| = \|d_k + \tilde{d}_k\| \le \|d_k\| + \|\tilde{d}_k\| \le \|d_k\| + 2\|\hat{d}_k\| \le (1 + 2c_1)\|d_k\|. \tag{3.19}$$

It then follows from (3.1), (3.2), (3.13) and (3.19) that

$$\begin{aligned}
\sum_{k\in T_2} \big| \|J_k^T F_k\| - \|J_{k+1}^T F_{k+1}\| \big| &= \sum_{k\in T_2} \big| (\|J_k^T F_k\| - \|J_k^T F_{k+1}\|) - (\|J_{k+1}^T F_{k+1}\| - \|J_k^T F_{k+1}\|) \big| \\
&\le \sum_{k\in T_2} (L_2^2 \|s_k\| + L_1 \|F_{k+1}\| \|s_k\|) \\
&\le \sum_{k\in T_2} (1 + 2c_1)(L_2^2 + L_1 \|F_0\|) \|d_k\| \\
&< +\infty.
\end{aligned} \tag{3.20}$$

Since (3.10) holds for infinitely many k, there exists a sufficiently large $\bar{k} \in T_2$ such that $\|J_{\bar{k}}^T F_{\bar{k}}\| \ge \varepsilon$ and

$$\sum_{k\in T_2, k\ge \bar{k}} \big| \|J_k^T F_k\| - \|J_{k+1}^T F_{k+1}\| \big| < \frac{\varepsilon}{2}. \tag{3.21}$$

By the induction principle, we obtain that $\|J_k^T F_k\| \ge \dfrac{\varepsilon}{2}$ and $k \in T_2$ or $x_{k+1} = x_k$ holds for all $k \ge \bar{k}$. Thus we have from (3.14) and (3.15) that

$$\|d_k\| \to 0, \text{ and } \ \lambda_k \to +\infty. \tag{3.22}$$

Now it follows from Lemma 3.3 that

$$\mu_k = \frac{\lambda_k}{\Lambda_k} = \frac{\lambda_k}{O(\|F_k\|^\delta)} \geq \frac{\lambda_k}{O(\|F_0\|^\delta)} \to +\infty. \tag{3.23}$$

By the relations (2.13), (2.25) and (2.26), we immediately have

$$\begin{aligned} |r_k - 1| &= \Big|\frac{Ared_k - Pred_k}{Pred_k}\Big| \\ &\leq \frac{\Big|\|F(x_k+s_k)\|^2 - \|F(y_k) + J_k\tilde{d}_k\|^2 + \|F(y_k)\|^2 - \|F_k + J_kd_k\|^2\Big|}{\|J_k^TF_k\| \min\Big\{\|d_k\|, \dfrac{\|J_k^TF_k\|}{\|J_k^TJ_k\|}\Big\}}. \end{aligned} \tag{3.24}$$

Now we turn our attention to the estimation of the numerator of (3.24). Since

$$\|F(y_k) - F_k - J_kd_k\| \leq L_1\|d_k\|^2, \ \|F(x_k+s_k) - F(y_k) - J_k\tilde{d}_k\| \leq L_1\|\tilde{d}_k\|^2 \tag{3.25}$$

and

$$\|F(y_k) + J_k\tilde{d}_k\| = \big\|\big[F(y_k) - F_k - J_kd_k\big] + F_k + J_ks_k\big\| \leq L_1\|d_k\|^2 + \|F_k + J_ks_k\|, \tag{3.26}$$

we have

$$\begin{aligned} &\big|\|F(y_k)\|^2 - \|F_k + J_kd_k\|^2\big| = \big|\|F(y_k)\| - \|F_k + J_kd_k\|\big|\big(\|F(y_k)\| + \|F_k + J_kd_k\|\big) \\ &\leq \big|\|F(y_k)\| - \|F_k + J_kd_k\|\big|\Big[\|F(y_k)\| - \|F_k + J_kd_k\| + 2\|F_k + J_kd_k\|\Big] \\ &\leq L_1\|d_k\|^2\big[L_1\|d_k\|^2 + 2\|F_k + J_kd_k\|\big] \\ &\leq O(\|d_k\|^2) \end{aligned} \tag{3.27}$$

and

$$\begin{aligned} &\big|\|F(x_k+s_k)\|^2 - \|F(y_k)+J_k\tilde{d}_k\|^2\big| \\ &= \big|\|F(x_k+s_k)\| - \|F(y_k) + J_k\tilde{d}_k\|\big|\big(\|F(x_k+s_k)\| + \|F(y_k) + J_k\tilde{d}_k\|\big) \\ &\leq \big|\|F(x_k+s_k)\| - \|F(y_k) + J_k\tilde{d}_k\|\big|\Big[\|F(x_k+s_k)\| - \|F(y_k) + J_k\tilde{d}_k\| + 2\|F(y_k) + J_k\tilde{d}_k\|\Big] \\ &\leq L_1\|\tilde{d}_k\|^2\big[L_1\|\tilde{d}_k\|^2 + 2\big\|F(y_k) + J_k\tilde{d}_k)\big\|\big] \\ &\leq L_1\|\tilde{d}_k\|^2\big[L_1\|\tilde{d}_k\|^2 + 2L_1\|d_k\|^2 + 2\|F_k + J_ks_k\|\big] \\ &\leq O(\|d_k\|^2). \end{aligned} \tag{3.28}$$

Substituting (3.27) and (3.28) into (3.24) yields

$$\begin{aligned} |r_k - 1| &= \Big|\frac{Ared_k - Pred_k}{Pred_k}\Big| \leq \frac{O(\|d_k\|^2)}{\|J_k^TF_k\| \min\Big\{\|d_k\|, \dfrac{\|J_k^TF_k\|}{\|J_k^TJ_k\|}\Big\}} \\ &\leq \frac{O(\|d_k\|^2)}{\dfrac{\varepsilon}{2}\min\Big\{\|d_k\|, \dfrac{\varepsilon}{2L_2^2}\Big\}} \leq \frac{O(\|d_k\|^2)}{O(\|d_k\|)} \to 0. \end{aligned} \tag{3.29}$$

This implies that $r_k \to 1$. Hence there exists a positive constant $\bar{\mu} > \mu$ such that $\mu_k < \bar{\mu}$ holds for sufficiently large k, which contradicts to (3.23).

Case 2: T_2 is finite. Then the set

$$T_3 = \left\{k : \|J_k^T F_k\| \geq \varepsilon \text{ and } x_{k+1} \neq x_k\right\}$$

is also finite. Denote $\tilde{k}$ be the largest index in T_3. We see that as long as $k \in T_1$ and $k > \tilde{k}$, then $x_{k+1} = x_k$. Define the set

$$T_4 = \left\{k : \|J_k^T F_k\| \geq \varepsilon \text{ and } x_{k+1} = x_k\right\}.$$

If $k \in T_4$, we can check that $\|J_{k+1}^T F_{k+1}\| \geq \varepsilon$ and $x_{k+2} = x_{k+1}$. This means that $k+1 \in T_4$. Now we deduce that $\|J_k^T F_k\| \geq \varepsilon$ and $x_{k+1} = x_k$ for all $k > \tilde{k}$. By Step 4 of Algorithm 2.1, it gives that $r_k < p_0 \leq p_1$. According to the updating rule of μ_k, we have

$$\mu_k \to +\infty. \tag{3.30}$$

Since $\lambda_k = \mu_k \Lambda_k$, it follows from (3.6) and (3.11) that

$$\lambda_k = \mu_k \Lambda_k \geq \mu_k \|F_{k+1}\|^\delta \geq \mu_k (L_2/\varepsilon)^\delta \to +\infty.$$

This means that

$$\|d_k\| = \| - (J_k^T J_k + \lambda_k I)^{-1} J_k^T F_k\| \to 0. \tag{3.31}$$

From (3.16)-(3.19), we have $r_k \to 1$ by using the same analysis as in (3.24)-(3.29). Hence there exists a positive constant $\bar{\mu} > \mu$ such that $\mu_k < \bar{\mu}$ holds for sufficiently large k, which is in contradiction with (3.30). This obtains the desired result. □

4 Local Convergence Rate of Algorithm 2.1

In this section, we will analyze the convergence rate of Algorithm 2.1 by using the singular value decomposition (SVD) technique. We assume that the sequence generated by Algorithm 2.1 converges to the solution set X^* and lies in some neighbourhood of $x^* \in X^*$.

We first give the following assumptions for obtaining the cubic convergence of Algorithm 2.1.

Assumption 4.1. (1) $F(x)$ is continuously differentiable, and both $F(x)$ and $J(x)$ are Lipschitz continuous on $N(x^*, b_1) = \{x : \|x - x^*\| \leq b_1\}$ with $b_1 \in (0, 1)$, i.e., there exist positive constants L_1 and L_2 such that

$$\|J(x) - J(y)\| \leq L_1 \|x - y\|, \ \forall x, y \in N(x^*, b_1) \tag{4.1}$$

and

$$\|F(x) - F(y)\| \leq L_2 \|x - y\|, \ \forall x, y \in N(x^*, b_1). \tag{4.2}$$

(2) $\|F(x)\|$ provide a local error bound on $N(x^*, b_1)$ for (1.1), i.e., there exists a positive constant $c > 0$ such that

$$\|F(x)\| \geq c\text{dist}(x, X^*), \quad \forall x \in N(x^*, b_1). \tag{4.3}$$

By (4.1) and (4.2), we obtain

$$\|F(y) - F(x) - J(x)(y - x)\| \leq L_1\|y - x\|^2, \ \forall x, y \in N(x^*, b_1). \tag{4.4}$$

According to the result derived by Behling and Iusem [5], there exists a positive number $\omega > 0$ such that

$$\operatorname{rank}(J(\bar{x})) = \operatorname{rank}(J(x^*)), \ \forall \bar{x} \in N(x^*, \omega) \cap X^* \tag{4.5}$$

when $F(x)$ provides a local error bound. Let $b \in (0, 1)$ and $b_1 = \min\{\omega, b\}$. Without loss of generality, we further assume that $x_k, y_k \in N(x^*, b_1/2)$. In the sequel, we let $\bar{x}_k$ be a vector in X^* such that $\|\bar{x}_k - x_k\| = \operatorname{dist}(x_k, X^*)$. Direct calculations give

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

Thus $\bar{x}_k \in N(x^*, b_1)$.

$\boxed{4.1}$ Properties of steps $\hat{d}_k$ and $\tilde{d}_k$

In this subsection, we investigate the property of $\hat{d}_k$ (or $\tilde{d}_k$) and give the relationship between the norm of $\hat{d}_k$ (or $\tilde{d}_k$) and the distance from x_k to the solution set.

Lemma 4.1. *Let* Assumptions 3.1 and 4.1 *hold. Then*

$$\|\hat{d}_k\| \leq O\big(\|\bar{x}_k - x_k\|\big) \quad and \quad \|\tilde{d}_k\| \leq O\big(\|\bar{x}_k - x_k\|\big) \tag{4.6}$$

hold for sufficiently large k.

Proof. Since $\delta \in [1, 2]$ and d_k is the minimizer of $\varphi_{k,1}(d) = \|F_k + J_k d\|^2 + \lambda_k\|d\|^2$, by the relations (2.5) and (3.3), it follows that

$$\begin{aligned}
\|d_k\|^2 &\leq \frac{\varphi_{k,1}(d_k)}{\lambda_k} \leq \frac{\varphi_{k,1}(\bar{x}_k - x_k)}{\lambda_k} \\
&= \frac{1}{\mu_k \Lambda_k}\left[\|F_k + J_k(\bar{x}_k - x_k)\|^2 + \mu_k\Lambda_k\|\bar{x}_k - x_k\|^2\right] \\
&= \frac{1}{\mu_k \Lambda_k}\|F_k + J_k(\bar{x}_k - x_k)\|^2 + \|\bar{x}_k - x_k\|^2 \\
&\leq \frac{1}{\mu O(\|F_k\|^\delta)}\|F_k + J_k(\bar{x}_k - x_k)\|^2 + \|\bar{x}_k - x_k\|^2 \\
&\leq \frac{L_1^2\|\bar{x}_k - x_k\|^4}{O(\|\bar{x}_k - x_k\|^\delta)} + \|\bar{x}_k - x_k\|^2 \leq O(\|\bar{x}_k - x_k\|^2).
\end{aligned} \tag{4.7}$$

By using the same method as in (3.16), we have

$$\|\hat{d}_k\| \leq L_1\|(J_k^T J_k + \lambda_k I)^{-1} J_k^T\|\|d_k\|^2 + 2\|d_k\|. \tag{4.8}$$

Now we turn our attention to the calculation of $\|(J_k^T J_k + \lambda_k I)^{-1} J_k^T\|$. According to the relation (4.5), we assume that $\operatorname{rank}(J(\bar{x})) = r$ for all $\bar{x} \in N(x^*, b_1) \cap X^*$. Correspondingly, the SVD of $J(\bar{x}_k)$ has the following form

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

where $\bar{\Sigma}_{k,1} = \mathrm{diag}(\bar{\sigma}_{k,1}, \bar{\sigma}_{k,2}, \cdots, \bar{\sigma}_{k,r})$ and $\bar{\sigma}_{k,1} \geq \bar{\sigma}_{k,2} \geq \cdots \geq \bar{\sigma}_{k,r} > 0$. Then the SVD of $J(x_k)$ is given by

$$\begin{aligned} J(x_k) = U_k \Sigma_k V_k^T &= (U_{k,1}, U_{k,2}, U_{k,3}) \begin{pmatrix} \Sigma_{k,1} & & \\ & \Sigma_{k,2} & \\ & & 0 \end{pmatrix} \begin{pmatrix} V_{k,1}^T \\ V_{k,2}^T \\ V_{k,3}^T \end{pmatrix} \\ &= U_{k,1}\Sigma_{k,1}V_{k,1}^T + U_{k,2}\Sigma_{k,2}V_{k,2}^T, \end{aligned}$$

where $\Sigma_{k,1} = \mathrm{diag}(\sigma_{k,1}, \sigma_{k,2}, \cdots, \sigma_{k,r})$, $\Sigma_{k,2} = \mathrm{diag}(\sigma_{k,r+1}, \sigma_{k,r+2}, \cdots, \sigma_{k,r+q})$, and $\sigma_{k,1} \geq \sigma_{k,2} \geq \cdots \geq \sigma_{k,r} > 0$, $\sigma_{k,r+1} \geq \sigma_{k,r+2} \geq \cdots \geq \sigma_{k,r+q} > 0$. We will neglect the subscript k if the context is clear in the sequel. Obviously the above formulas can be written as

$$J_k = U_1\Sigma_1V_1^T + U_2\Sigma_2V_2^T.$$

Direct calculations give

$$\begin{aligned} &\|(J_k^TJ_k + \lambda_k I)^{-1}J_k^T\| \\ &= \left\| (V_1, V_2, V_3) \begin{pmatrix} (\Sigma_1^2 + \lambda_k I)^{-1}\Sigma_1 & & \\ & (\Sigma_2^2 + \lambda_k I)^{-1}\Sigma_2 & \\ & & 0 \end{pmatrix} \begin{pmatrix} U_1^T \\ U_2^T \\ U_3^T \end{pmatrix} \right\| \\ &\leq \left\|(\Sigma_1^2 + \lambda_k I)^{-1}\Sigma_1\right\| + \left\|\lambda_k^{-1}\Sigma_2\right\|. \end{aligned} \tag{4.9}$$

According to the Lipschitzness of J_k and matrix perturbation theory [32], it follows that

$$\|\mathrm{diag}(\Sigma_1 - \bar{\Sigma}_1, \Sigma_2, 0)\| \leq \|J_k - \bar{J}_k\| \leq L_1\|\bar{x}_k - x_k\|.$$

This means that

$$\|\Sigma_1 - \bar{\Sigma}_1\| \leq L_1\|\bar{x}_k - x_k\| \quad \text{and} \quad \|\Sigma_2\| \leq L_1\|\bar{x}_k - x_k\|. \tag{4.10}$$

Correspondingly

$$\left\|\lambda_k^{-1}\Sigma_2\right\| = \frac{\|\Sigma_2\|}{\mu_k\Lambda_k} \leq \frac{\|\Sigma_2\|}{\mu\Lambda_k} = \frac{\|\Sigma_2\|}{\mu O(\|F_k\|^\delta)} = \frac{L_1\|\bar{x}_k - x_k\|}{O(\|\bar{x}_k - x_k\|^\delta)} = O(\|\bar{x}_k - x_k\|^{1-\delta}). \tag{4.11}$$

Note that

$$\frac{\sigma_i}{\sigma_i^2 + \lambda_k} \leq \frac{\sigma_i}{2\sigma_i\sqrt{\lambda_k}} = \frac{1}{2\sqrt{\lambda_k}}$$

for $\sigma_i > 0$ and $i = 1, 2, \cdots, r$. This gives that

$$\left\|(\Sigma_1^2 + \lambda_k I)^{-1}\Sigma_1\right\| \leq \frac{1}{2\sqrt{\mu_k\Lambda_k}} \leq \frac{1}{2\sqrt{\mu O(\|F_k\|^\delta)}} \leq O(\|\bar{x}_k - x_k\|^{-\frac{\delta}{2}}). \tag{4.12}$$

Combining (4.9), (4.11) and (4.12), we have

$$\|(J_k^TJ_k + \lambda_k I)^{-1}J_k^T\| \leq \left\|(\Sigma_1^2 + \lambda_k I)^{-1}\Sigma_1\right\| + \left\|\lambda_k^{-1}\Sigma_2\right\| \leq O(\|\bar{x}_k - x_k\|^{-\frac{\delta}{2}}).$$

This together with (4.7) and (4.8), yields

$$\|\hat{d}_k\| \leq L_1 O\big(\|\bar{x}_k - x_k\|^{-\frac{\delta}{2}}\big)\|d_k\|^2 + 2\|d_k\| \leq O\big(\|\bar{x}_k - x_k\|\big). \tag{4.13}$$

It then follows from (3.18) that

$$\|\tilde{d}_k\| \leq 2\|\hat{d}_k\| \leq O\big(\|\bar{x}_k - x_k\|\big). \tag{4.14}$$

This gives the desired result. □

$\boxed{4.2}$ Boundedness of the LM parameter

From the updating rule of $\{\mu_k\}$, it follows that $\{\mu_k\}$ is bounded below. In this subsection, we prove that $\{\mu_k\}$ is bounded above.

Lemma 4.2. *Let* Assumptions 3.1 and 4.1 *hold. Then* $\mu_k \le \bar{\mu}$ *holds for all sufficiently large* k*, where* $\bar{\mu}$ *is a positive number such that* $\bar{\mu} > \mu$*.*

Proof. We first show that the following two inequalities

$$\|F_k\|^2 - \|F_k + J_k d_k\|^2 \ge \check{c}\|F_k\| \min\left\{\|d_k\|, \|\bar{x}_k - x_k\|\right\} \tag{4.15}$$

and

$$\|F(y_k)\|^2 - \|F(y_k) + J_k \hat{d}_k\|^2 \ge \tilde{c}\|F(y_k)\| \min\left\{\|\hat{d}_k\|, \|\bar{y}_k - y_k\|\right\} \tag{4.16}$$

hold for sufficiently large k, where $\check{c}$ and $\tilde{c}$ are positive constants.

For proving (4.15), we divide two cases. For the case that $\|\bar{x}_k - x_k\| \le \|d_k\|$. Since d_k is the minimizer of $\varphi_{k,1}$, it follows from the relations (4.3) and (4.4) that

$$\|F_k\| - \|F_k + J_k d_k\| \ge \|F_k\| - \|F_k + J_k(\bar{x}_k - x_k)\| \ge c\|\bar{x}_k - x_k\| - L_1\|\bar{x}_k - x_k\|^2 \ge \check{c}\|\bar{x}_k - x_k\|. \tag{4.17}$$

For the case that $\|\bar{x}_k - x_k\| > \|d_k\|$. Similar to above, we have

$$\begin{aligned}
\|F_k\| - \|F_k + J_k d_k\| &\ge \|F_k\| - \left\|F_k + \frac{\|d_k\|}{\|\bar{x}_k - x_k\|} J_k(\bar{x}_k - x_k)\right\| \\
&\ge \frac{\|d_k\|}{\|\bar{x}_k - x_k\|}(\|F_k\| - \|F_k + J_k(\bar{x}_k - x_k)\|) \\
&\ge \frac{\|d_k\|}{\|\bar{x}_k - x_k\|}(c\|\bar{x}_k - x_k\| - L_1\|\bar{x}_k - x_k\|^2) \ge \check{c}\|d_k\|. \qquad (4.18)
\end{aligned}$$

Combining (4.17) with (4.18) gives

$$\begin{aligned}
\|F_k\|^2 - \|F_k + J_k d_k\|^2 &\ge (\|F_k\| + \|F_k + J_k d_k\|)(\|F_k\| - \|F_k + J_k d_k\|) \\
&\ge \check{c}\|F_k\| \min\left\{\|d_k\|, \|\bar{x}_k - x_k\|\right\},
\end{aligned}$$

which yields (4.15).

Now we prove (4.16). If $\|\bar{y}_k - y_k\| \le \|\hat{d}_k\|$, by the relations (4.2)-(4.4) and the fact that $\hat{d}_k$ is the minimizer of $\varphi_{k,2}$, we have

$$\begin{aligned}
\|F(y_k)\| - \|F(y_k) + J_k\hat{d}_k\| &\ge \|F(y_k)\| - \|F(y_k) + J_k(\bar{y}_k - y_k)\| \\
&\ge \|F(y_k)\| - \|F(y_k) + J(y_k)(\bar{y}_k - y_k)\| \\
&\quad - \|J_k - J(y_k)\|\|\bar{y}_k - y_k\| \\
&\ge c\|\bar{y}_k - y_k\| - L_1\|\bar{y}_k - y_k\|^2 - L_1\|d_k\|\|\bar{y}_k - y_k\| \\
&\ge \tilde{c}\|\bar{y}_k - y_k\|. \qquad (4.19)
\end{aligned}$$

If $\|\bar{y}_k - y_k\| > \|\hat{d}_k\|$, we get

$$\begin{aligned}
\|F(y_k)\| - \|F(y_k) + J_k\hat{d}_k\| &\ge \|F(y_k)\| - \left\|F(y_k) + \frac{\|\hat{d}_k\|}{\|\bar{y}_k - y_k\|}J_k(\bar{y}_k - y_k)\right\| \\
&\ge \frac{\|\hat{d}_k\|}{\|\bar{y}_k - y_k\|}(\|F(y_k)\| - \|F(y_k) + J_k(\bar{y}_k - y_k)\|) \\
&\ge \frac{\|\hat{d}_k\|}{\|\bar{y}_k - y_k\|}\bar{c}\|\bar{y}_k - y_k\| \ge \tilde{c}\|\hat{d}_k\|. \qquad (4.20)
\end{aligned}$$

Combining (4.19) with (4.20) gives

$$\begin{aligned}\|F(y_k)\|^2 - \|F(y_k) + J_k\hat{d}_k\|^2 &\geq (\|F(y_k)\| + \|F(y_k) + J_k\hat{d}_k\|)(\|F(y_k)\| - \|F(y_k) + J_k\hat{d}_k\|)\\ &\geq \tilde{c}\|F(y_k)\| \min\left\{\|\hat{d}_k\|, \|\bar{y}_k - y_k\|\right\},\end{aligned}$$

which yields (4.16).

Next we show $r_k \to 1$. Since d_k is the minimizer of $\varphi_{k,1}$, by the relations (4.2) and (4.13), we have

$$\|F_k + J_k d_k\| \leq \|F_k\| \leq L_2\|\bar{x}_k - x_k\|$$

and

$$\|F_k + J_k s_k\| \leq \|F_k + J_k d_k\| + \|J_k\hat{d}_k\| \leq L_2\|\bar{x}_k - x_k\| + L_2\|\hat{d}_k\| \leq \check{c}_2\|\bar{x}_k - x_k\|.$$

It then follows from (3.18), (3.27), (3.28), (4.13) and Lemma 4.1 that

$$\big|\|F(y_k)\|^2 - \|F_k + J_k d_k\|^2\big| \leq L_1\|d_k\|^2\big[L_1\|d_k\|^2 + 2\|F_k + J_k d_k\|\big] \leq O(\|d_k\|^2\|\bar{x}_k - x_k\|)$$

and

$$\begin{aligned}&\big|\|F(x_k + s_k)\|^2 - \|F(y_k) + J_k\tilde{d}_k\|^2\big| \leq L_1\|\tilde{d}_k\|^2\big[L_1\|\tilde{d}_k\|^2 + 2\big\|F(y_k) + J_k\tilde{d}_k\big\|\big]\\ \leq\ & L_1\|\tilde{d}_k\|^2\big[L_1\|\tilde{d}_k\|^2 + 2L_1\|d_k\|^2 + 2\|F_k + J_k s_k\|\big]\\ \leq\ & O(\|d_k\|^2\|\bar{x}_k - x_k\|).\end{aligned}$$

The above two inequalities imply that

$$|Ared_k - Pred_k| \leq O(\|d_k\|^2\|\bar{x}_k - x_k\|). \tag{4.21}$$

From (2.23), (2.25), (4.15) and (4.16), we obtain

$$\begin{aligned}Pred_k &= \|F_k\|^2 - \|F_k + J_k d_k\|^2 + \|F(y_k)\|^2 - \|F(y_k) + J_k\tilde{d}_k\|^2\\ &= \|F_k\|^2 - \|F_k + J_k d_k\|^2 + \|F(y_k)\|^2 - \|F(y_k) + J_k\hat{d}_k\|^2\\ &\quad + \|F(y_k) + J_k\hat{d}_k\|^2 - \|F(y_k) + J_k\tilde{d}_k\|^2\\ &\geq \|F_k\|^2 - \|F_k + J_k d_k\|^2 + \|F(y_k)\|^2 - \|F(y_k) + J_k\hat{d}_k\|^2\\ &\geq O(\|d_k\|\|\bar{x}_k - x_k\|),\end{aligned} \tag{4.22}$$

which, together with (4.21), yields

$$|r_k - 1| \quad = \quad \Big|\frac{Ared_k - Pred_k}{Pred_k}\Big| \leq \frac{O(\|d_k\|^2\|\bar{x}_k - x_k\|)}{O(\|d_k\|\|\bar{x}_k - x_k\|)} \to 0. \tag{4.23}$$

Thus $r_k \to 1$, and then there exists a positive constant $\bar{\mu}$ such that $\mu_k < \bar{\mu}$ holds for sufficiently large k. This gives the desired result. □

4.3 Convergence order of Algorithm 2.1

To obtain the convergence order of Algorithm 2.1, we need the following two lemmas.

Lemma 4.3 ([13]). *Let* Assumptions 3.1 and 4.1 *hold. If* $x_k \in N(x^*, b_1/2)$, *then*
(1) $\|U_1U_1^TF_k\| \leq L_2\|\bar{x}_k - x_k\|$;
(2) $\|U_2U_2^TF_k\| \leq 3L_1\|\bar{x}_k - x_k\|^2$;
(3) $\|U_3U_3^TF_k\| \leq L_1\|\bar{x}_k - x_k\|^2$.

Lemma 4.4 ([6]). *Let* Assumptions 3.1 and 4.1 *hold. If* $x_k \in N(x^*, b_1/2)$, *then*
(1) $\|U_1U_1^T F(y_k)\| \leq O(\|\bar{x}_k - x_k\|^2)$;
(2) $\|U_2U_2^T F(y_k)\| \leq O(\|\bar{x}_k - x_k\|^3)$;
(3) $\|U_3U_3^T F(y_k)\| \leq O(\|\bar{x}_k - x_k\|^3)$.

In what follows, we will show the cubic convergence of Algorithm 2.1.

Theorem 4.5. *The convergence rate of* Algorithm 2.1 *is of order* 3 *under the conditions of* Assumptions 3.1 and 4.1.

Proof. Since $x_k \in N(x^*, b_1/2)$, we may assume that $L_1\|\bar{x}_k - x_k\| \leq \bar{\sigma_r}/2$ for sufficiently large k. According to (4.10), it follows that

$$|\bar{\sigma_r} - \sigma_r| \leq L_1\|\bar{x}_k - x_k\| \leq \bar{\sigma_r}/2.$$

Correspondingly

$$\|(\Sigma_1^2 + \lambda_k I)^{-1}\| \leq \|\Sigma_1^{-2}\| = \left|\frac{1}{\sigma_r}\right|^2 \leq \left|\frac{1}{\bar{\sigma_r} - L_1\|\bar{x}_k - x_k\|}\right|^2 \leq \frac{4}{\bar{\sigma}_r^2} \tag{4.24}$$

and

$$\|(\Sigma_1^2 + \lambda_k I)^{-2}\| \leq \|\Sigma_1^{-4}\| \leq \left|\frac{1}{\sigma_r}\right|^4 \leq \left|\frac{1}{\bar{\sigma_r} - L_1\|\bar{x}_k - x_k\|}\right|^4 \leq \frac{16}{\bar{\sigma}_r^4} \tag{4.25}$$

hold for sufficiently large k. Since $\delta \in [1, 2]$, we have from (4.11), (4.12), (4.24) and Lemma 4.4 that

$$\begin{aligned}
\|\hat{d}_k\| &= \| - V_1(\Sigma_1^2 + \lambda_k I)^{-1}\Sigma_1 U_1^T F(y_k) - V_2(\Sigma_2^2 + \lambda_k I)^{-1}\Sigma_2 U_2^T F(y_k)\| \\
&\leq \|\Sigma_1^{-1}\|\|U_1^T F(y_k)\| + \|\lambda_k^{-1}\Sigma_2\|\|U_2^T F(y_k)\| \\
&\leq O(\|\bar{x}_k - x_k\|^2) + O(\|\bar{x}_k - x_k\|^{4-\delta}) \\
&\leq O(\|\bar{x}_k - x_k\|^2).
\end{aligned} \tag{4.26}$$

This together with (3.18) yields

$$\|\tilde{d}_k\| \leq 2\|\hat{d}_k\| \leq O(\|\bar{x}_k - x_k\|^2). \tag{4.27}$$

By (2.8), (2.21) and some direct calculations, we have

$$\begin{aligned}
&F(y_k) + J_k\tilde{d}_k \\
&= F(y_k) + J_k(\hat{d}_k + d_k^c) = F(y_k) + J_k\hat{d}_k + J_k(\lambda_k(J_k^T J_k + \lambda_k I)^{-1}\hat{d}_k) \\
&= F(y_k) - J_k(J_k^T J_k + \lambda_k I)^{-1}J_k^T F(y_k) - \lambda_k J_k(J_k^T J_k + \lambda_k I)^{-2}J_k^T F(y_k) \\
&= [I - J_k(J_k^T J_k + \lambda_k I)^{-1}J_k^T - \lambda_k J_k(J_k^T J_k + \lambda_k I)^{-2}J_k^T]F(y_k) \\
&= \begin{pmatrix} U_1, & U_2, & U_3 \end{pmatrix} \begin{pmatrix} \lambda_k^2(\Sigma_1^2 + \lambda_k I)^{-2} & & \\ & \lambda_k^2(\Sigma_2^2 + \lambda_k I)^{-2} & \\ & & I \end{pmatrix} \begin{pmatrix} U_1^T \\ U_2^T \\ U_3^T \end{pmatrix} F(y_k) \\
&= \lambda_k^2 U_1(\Sigma_1^2+\lambda_k I)^{-2}U_1^T F(y_k) + \lambda_k^2 U_2(\Sigma_2^2+\lambda_k I)^{-2}U_2^T F(y_k) + U_3U_3^T F(y_k).
\end{aligned} \tag{4.28}$$

Notice that LM parameter satisfies

$$\lambda_k = \mu_k\Lambda_k = \mu_k O(\|F_k\|^\delta) \leq \bar{\mu}O(\|F_k\|^\delta) \leq O(\|\bar{x}_k - x_k\|^\delta),\ \delta \in [1, 2] \tag{4.29}$$

and
$$\|(\Sigma_2^2+\lambda_k I)^{-2}\| \le \frac{1}{\lambda_k^2}. \tag{4.30}$$
By using (4.25), (4.28)-(4.30) and Lemma 4.4, we immediately have
$$\begin{aligned}\|F(y_k)+J_k\tilde{d}_k\| &\le O(\|\bar{x}_k-x_k\|^{2\delta})O(\|\bar{x}_k-x_k\|^2)+O(\|\bar{x}_k-x_k\|^3)+O(\|\bar{x}_k-x_k\|^3)\\ &= O(\|\bar{x}_k-x_k\|^3). \end{aligned}\tag{4.31}$$
According to the relations (4.1), (4.3) and (4.4), we obtain
$$\begin{aligned} c\|\bar{x}_{k+1}-x_{k+1}\| &\le \|F(x_{k+1})\|=\|F(x_k+s_k)\|=\|F(y_k+\tilde{d}_k)\|\\ &\le \|F(y_k)+J(y_k)\tilde{d}_k\|+L_1\|\tilde{d}_k\|^2\\ &\le \|F(y_k)+J_k\tilde{d}_k\|+\|(J(y_k)-J_k)\tilde{d}_k\|+L_1\|\tilde{d}_k\|^2\\ &\le \|F(y_k)+J_k\tilde{d}_k\|+L_1\|d_k\|\|\tilde{d}_k\|+L_1\|\tilde{d}_k\|^2. \end{aligned}$$
It then follows from the relations (4.27) and (4.31) and Lemma 4.1 that
$$c\|\bar{x}_{k+1}-x_{k+1}\| \le O(\|\bar{x}_k-x_k\|^3)+O(\|\bar{x}_k-x_k\|^3)+O(\|\bar{x}_k-x_k\|^4) \le O(\|\bar{x}_k-x_k\|^3). \tag{4.32}$$
This means that the sequence $\{x_k\}$ generated by Algorithm 2.1 converges to the solution set X^* with 3th order.

Since
$$\|\bar{x}_k-x_k\|=\mathrm{dist}(x_k,X^*) \le \|\bar{x}_{k+1}-x_k\| \le \|\bar{x}_{k+1}-x_{k+1}\|+\|s_k\|,$$
we have from the relation (4.32) that
$$\|\bar{x}_k-x_k\| \le 2\|s_k\|$$
hold for sufficiently large k. By the relation (4.32) and Lemma 4.1, we have
$$\|s_{k+1}\| \le O(\|s_k\|^3).$$
This means that the sequence $\{x_k\}$ converges to some solution $x^* \in X^*$ with 3th order. This proves the desired result. □

5 Numerical Experiments

In this section, we conduct several experiments to verify the effectiveness of Algorithm 2.1. All of the tests are run in MATLAB R2015a with the machine precision 10^{-16} on a personal computer (Intel (R) Core (TM)i7-5500U), where the CPU is 2.40 GHz and the memory is 8.0 GB.

We compare our proposed algorithm with the modified two steps LM method (denoted by MTLM method) developed by Amini and Rostami [3], the accelerated modified LM method (denoted by AMLM method) developed by Fan [14], and the LM algorithm with correction (denoted by LMC method) developed by Fan and Zeng [19]. Algorithm 2.1 is divided to NLMC1 and NLMC2 with $\delta=1$ and $\delta=2$, respectively. In addition, if we delete the correction step d_k^c, i.e., $s_k=d_k+\hat{d}_k$, which results the nonmonotone LM method without correction, abbreviated as NLM1 for $\delta=1$ and NLM2 for $\delta=2$. Similarly, we test LMC (AMLM) algorithm with $\delta=1$ and $\delta=2$, denoted by LMC1 (AMLM1) and LMC2 (AMLM2), respectively.

The above mentioned algorithms are terminated when the norm of $J_k^T F_k$, i.e., the derivative of $\frac{1}{2}\|F(x)\|^2$ at the kth iteration, is less than $\varepsilon = 10^{-6}$ or k reaches the maximum number of iterations, e.g., 500.

We measure the effectiveness of the above mentioned iterative algorithms by the number of function calculations (NF) and the number of Jacobian calculations (NJ). We usually employ "NT=NF+n*NJ"to measure the total computation since the Jacobian calculation typically requires n times as much computations as function. In addition, the sign "-"indicates that the algorithms can not converge to a solution within 500 iterations.

5.1 Performance profile

In this subsection, we apply the performance profile of Dolan and Moŕe [10] to demonstrate the overall behaviour of the considered algorithms and obtain more insight about the performance of the considered algorithms. For the algorithm implementation, we choose $p_0 = 0.0001$, $p_1 = 0.25$, $p_2 = 0.75$, $\mu = 10^{-8}$ and $\mu_0 = 10^{-4}$. We also let $\eta = 0.75$ and $N = 10$.

We select test problems from a wide range of literatures with dimensions ranging from 10 to 1000. More concretely, the first four problems are selected from [28], the last five problems are chosen from [8] and the others are discussed in [26]. For all these algorithms, the Jacobian matrix J_k can be either computed analytically by a user-supplied function or approximated by using finite differences formula in the form of

$$[J_k]_{\cdot j} \sim \frac{1}{h_j}(F(x_k + h_j e_j) - F_k),$$

where e_j denotes the jth column of the identity matrix, $[J_k]_{\cdot j}$ is the jth column of J_k and

$$h_j = \begin{cases} \sqrt{\epsilon_m}, & \text{if } x_{k_j} = 0, \\ \sqrt{\epsilon_m}\text{sign}(x_{k_j}) \max\{|x_{k_j}|, \dfrac{\|x_k\|_1}{n}\}, & \text{otherwise.} \end{cases}$$

Here ϵ_m is the machine epsilon specified by the Matlab function "eps".

For the sake of completeness, we now give the notion of a performance profile as a means to evaluate and compare the performance of the solvers on a test set $\mathcal{P}$. Assume that we have n_s solvers and n_p problems. For each test problem p and solver s, we define

$$k_{p,s} = \text{iteration number required to solve problem } p \text{ by solver } s$$

and

$$c_{p,s} = \text{total calculations required to solve problem } p \text{ by solver } s.$$

Following the idea of Dolan and Moŕe [10], we compare the performance on problem p by solver s with the best performance by any solver on this problem; that is, we use the performance ratios

$$\rho^k_{p,s} = \frac{k_{p,s}}{\min\{k_{p,s} : 1 \le s \le n_s\}} \quad \text{and} \quad \rho^c_{p,s} = \frac{c_{p,s}}{\min\{c_{p,s} : 1 \le s \le n_s\}}.$$

When the solver s has failed on problem p, we set $\rho^k_{p,s} = r^k_{\text{fail}}$ and $\rho^c_{p,s} = r^c_{\text{fail}}$, where r^k_{fail} and ${}^c_{\text{fail}}$ are strictly larger than any performance ratio. For any factor $\tau \ge 1$, the overall performance of solver s is specified by

$$p^k_s(\tau) = \frac{1}{n_p}\text{size}\{p \in \mathcal{P} : \rho^k_{p,s} \le \tau\}$$

and

$$p_s^c(\tau) = \frac{1}{n_p}\text{size}\{p \in \mathcal{P} : \rho_{p,s}^c \leq \tau\}.$$

Hence performance profiles, for every $\tau \geq 1$, produce the proportions $p_s^k(\tau)$ and $p_s^c(\tau)$ of test problems on which each solver has a performance within a factor τ of the best. That is, we plot

$$\tau \mapsto \frac{1}{n_p}\text{size}\{p \in \mathcal{P} : \rho_{p,s}^k \leq \tau\} \text{ or } \tau \mapsto \frac{1}{n_p}\text{size}\{p \in \mathcal{P} : \rho_{p,s}^c \leq \tau\}$$

to measure the performance on problem p by solver s. It is easy to see that $p_s^k(1)$ (or $p_s^c(1)$) is the probability in which the solver s is the best and $\lim_{\tau \to r_{\text{fail}}^k} p_s^k(\tau)$ (or $\lim_{\tau \to r_{\text{fail}}^k} p_s^k(\tau)$) gives the ratio of test problems of $\mathcal{P}$ for which the algorithm s succeeded. Consequently, the values on the left side of the figures give the information about the efficiency of each solver and the values on the right side represent the robustness of the solvers. This implies that the best solver is the highest on the figures.

The performance of all algorithms, based on both the iteration number and total calculations (NT), have been, respectively, assessed in Figure 1 and Figure 2. It is seen from Figure 1 that NLMC1, NLM1, LMC1 [19], AMLM1 [14], NLMC2, NLM2, LMC2 [19], AMLM2 [14] and MTLM [3] are all feasible while NLMC1, NLM1 and LMC1 are better than NLMC2, NLM2 and LMC2. We clearly see from Figure 1 that NLMC1 win in nearly 76% of the test problems with the greatest efficiency. Furthermore, we see from Figure 2 that NLMC1 is better than others where it has most won in approximately 56% of the test problems concerning the total number of calculations.

We also report the NJ, NF and NT for each algorithm in Table 1. From Table 1, we find that in most cases the iteration number, i.e., the number of Jacobian calculations, of our proposed algorithm is less than the one in MTLM method [3]. It is known that our proposed nonmonotone LM method with correction (NLMC) is superior to the accelerated modified LM method (AMLM) in [14] and the LM method with correction (LMC) in [19] since it requires fewer computations for most test problems. We also observe from Table 1 that the correction technique is quite efficient and the NLMC algorithm outperforms the NLM algorithm for most of the test set problems.

Table 1: Numerical results for the test set

Problem name(Dim)	NLMC1 NF/NJ/NT	NLM1 NF/NJ/NT	LMC1 [19] NF/NJ/NT	AMLM1 [14] NF/NJ/NT	NLMC2 NF/NJ/NT	NLM2 NF/NJ/NT	LMC2 [19] NF/NJ/NT	AMLM2 [14] NF/NJ/NT	MTLM [3] NF/NJ/NT
Extend-Rosenbrock(40)	5/3/125	7/4/167	5/5/205	5/3/125	7/4/167	7/4/167	5/5/205	5/3/125	9/5/209
Distrete boundary(300)	7/4/1207	7/4/1207	3/3/903	9/5/1509	7/4/1207	7/4/1207	3/3/903	7/4/1207	11/6/1811
Discrete integral(1000)	7/4/4007	7/4/4007	5/5/5005	7/4/4007	7/4/4007	7/4/4007	5/5/5005	7/4/4007	7/4/4007
Watson(1000)	5/3/3005	5/3/3005	4/4/4004	5/3/3005	5/3/3005	7/4/4007	4/4/4004	5/3/3005	9/5/5009
Trigonometric function(1000)	77/20/20077	77/22/22077	70/48/48070	-	125/35/35125	111/31/31111	79/48/48079	65/16/16065	79/28/28079
Triexp function(10)	27/7/97	15/7/85	7/7/77	13/6/73	25/7/95	27/7/97	7/7/77	17/7/87	13/6/73
Chandrasekhar's-H(100)	17/9/917	17/9/917	16/16/1616	17/9/917	17/9/917	17/9/917	16/16/1616	17/9/917	17/9/917
Tridiagonal system(300)	113/34/10313	113/35/10613	87/48/14487	-	109/34/10309	125/43/13025	94/48/14494	115/36/10915	123/47/14223
Five-diagonal system(300)	103/26/7903	121/31/9421	60/35/10560	113/30/9113	101/29/8801	105/30/9105	58/35/10558	105/29/8805	117/33/10017
Seven-diagonal system(300)	21/7/2121	23/8/2423	19/13/3919	27/8/2427	17/7/2117	29/11/3329	16/13/3916	23/8/2423	21/9/2721
Exponential(300)	15/8/2415	15/8/2415	11/11/3311	21/11/3321	15/8/2415	15/8/2415	11/11/3311	21/11/3321	15/8/2415
Exponential 1(1000)	21/11/11021	21/11/11021	14/14/14014	23/12/12023	21/11/11021	21/11/11021	14/14/14014	23/12/12023	21/11/11021
Exponential 2(300)	9/5/1509	9/5/1509	6/6/1806	7/4/1207	11/6/1811	11/6/1811	6/6/1806	9/5/1509	11/6/1811
Function 18(300)	19/6/1819	19/6/1819	13/8/2413	19/6/1819	15/7/2115	15/7/2115	9/7/2109	17/7/2117	11/6/1811
Function 21(1000)	17/6/6017	17/6/6017	11/7/7011	17/6/6017	13/6/6013	13/6/6013	8/7/7008	15/6/6015	11/6/6011
Function 27(1000)	17/9/9017	13/7/7013	15/14/14015	31/12/12031	21/10/10021	19/9/9019	18/13/13018	23/12/12023	23/10/10023
Logarithmic(1000)	7/4/4007	7/4/4007	5/5/5005	7/4/4007	7/4/4007	7/4/4007	5/5/5005	7/4/4007	7/4/4007
Strictly convex 1(1000)	7/4/4007	7/4/4007	5/5/5005	7/4/4007	7/4/4007	7/4/4007	5/5/5005	7/4/4007	7/4/4007
Strictly convex 2(1000)	15/8/8015	15/8/8015	10/10/10010	-	17/9/9017	15/8/8015	10/10/10010	17/9/9017	15/8/8015
Complementarity(1000)	43/14/14043	31/13/13031	28/24/24028	29/12/12029	27/13/13027	27/14/14027	26/24/24026	27/13/13027	23/12/12023
Extend-Gragg-Levy(10)	13/7/83	13/7/83	9/9/99	19/10/119	13/7/83	13/7/83	9/9/99	19/10/119	13/7/83
Variably-dimension(30)	3/2/63	3/2/63	2/2/62	3/2/63	3/2/63	5/3/95	3/3/93	3/2/63	3/2/63
Zero Jacobian(1000)	5/3/3005	5/3/3005	5/5/5005	5/3/3005	5/3/3005	5/3/3005	7/7/7007	5/3/3005	5/3/3005
Generalized-broyden(1000)	23/12/12023	23/12/12023	22/22/22022	21/11/11021	23/12/12023	23/12/12023	22/22/22022	21/11/11021	23/12/12023
Extend-Powell-singular(1000)	15/8/8015	15/8/8015	10/10/10010	19/10/10019	17/9/9017	17/9/9017	10/10/10010	19/10/10019	15/8/8015

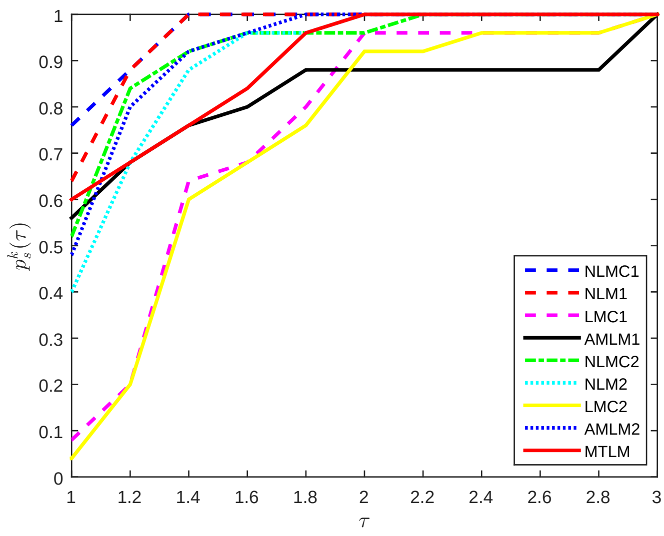

Figure 1: Performance profile for the number of iterations

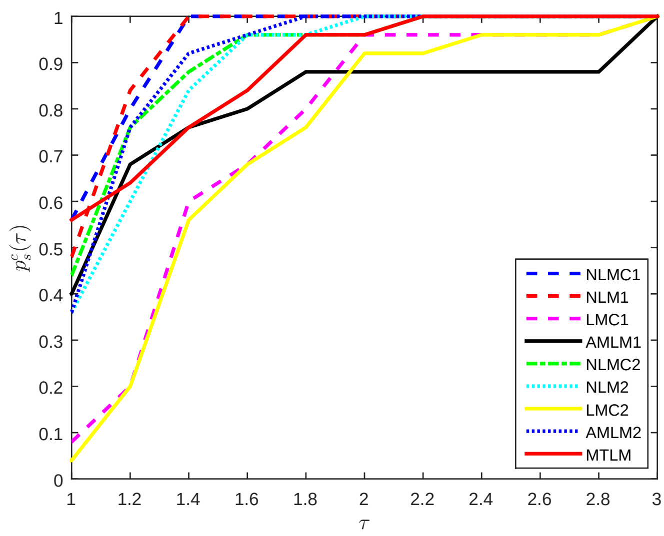

Figure 2: Performance profile for the total number of calculations

5.2 Singular problem

In this subsection, we test the algorithm on some singular problems, which were generated by modifying the nonsingular problems given by Moré et al. in [28], and have the same form as in [30]

$$\hat{F}(x) = F(x) - J(x^*)A(A^TA)^{-1}A^T(x - x^*), \tag{5.1}$$

where $F(x)$ is the standard nonsingular test function, $A \in \mathbb{R}^{n-k}$ $(1 \leq k \leq n)$ is a full column rank matrix, and x^* is the root of $F(x)$. It is easy to check that $\hat{F}(x^*) = 0$ and

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

has rank $n-k$. More concretely, we choose the rank of $\hat{J}(x^*)$ to be $n-1$ and $n-2$ respectively, by using

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

and

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

respectively.

The results for the first set of problems of $\mathrm{rank}(F'(x^*)) = n-1$ are summarized in Table 2, and the second set of problems of $\mathrm{rank}(F'(x^*)) = n-2$ are listed in Table 3. With regard to Tables 2 and 3, the second column of the table indicates that the start point is $0.01x_0$, $0.1x_0$, x_0, $10x_0$, $100x_0$, where x_0 is the initial value suggested by Moré et. al [28]; "Label" denotes the problem number in Moré et. al [28]; "n" indicates the dimension of test functions; "F" indicates a final value of the norm of $J_k^T F_k$.

Table 2: Numerical results on the first singular test set with $\mathrm{rank}(F'(x^*)) = n-1$.

Label(n)	x_0	NLMC1	LMC1[19]	AMLM1[14]	NLMC2	LMC2 [19]	AMLM2[14]	MTLM [3]
		NJ/NF/NT/F	NJ/NF/NT/F	NJ/NF/NT/F	NJ/NF/NT/F	NJ/NF/NT/F	NJ/NF/NT/F	NJ/NF/NT/F
2(2)	1	11/21/43/7.5367e-07	15/15/45/9.5515e-07	11/21/43/7.5041e-07	15/15/45/9.5515e-07	11/21/43/9.1716e-07	11/21/43/8.3307e-07	15/15/45/9.6037e-07
	10	17/33/67/1.5506e-07	23/23/69/3.2506e-07	17/33/67/1.5362e-07	18/35/71/2.1307e-07	23/23/69/4.3473e-07	17/33/67/3.1321e-07	17/33/67/1.5358e-07
	100	21/41/83/2.2309e-07	29/29/87/2.6563e-07	21/41/83/2.2371e-07	26/51/103/1.8553e-07	30/30/90/5.7491e-07	23/45/91/2.2635e-07	21/41/83/2.3041e-07
7(3)	1	6/11/29/4.5050e-09	8/8/32/1.1564e-11	7/13/34/4.1504e-12	7/13/34/5.4451e-14	10/10/40/6.1353e-13	7/13/34/7.3317e-08	14/27/69/1.8591e-07
	10	6/11/29/7.9691e-14	9/9/36/4.2625e-07	13/29/68/9.9058e-07	25/33/108/6.3083e-10	14/14/56/3.2485e-08	19/51/108/7.7344e-07	14/27/69/1.9874e-07
	100	13/25/64/5.7110e-07	15/15/60/8.3823e-07	14/27/69/2.3617e-07	8/15/39/4.4113e-08	16/16/64/5.1973e-08	15/29/74/1.1928e-09	8/15/39/7.6386e-09
14(4)	1	13/25/77/2.7554e-07	18/18/90/2.6584e-07	13/25/77/2.7542e-0707	13/25/77/3.3778e-07	18/18/90/2.8352e-07	13/25/77/3.5724e-07	13/25/77/2.9509e-07
	0.1	12/23/71/2.0749e-07	16/16/80/4.4925e-07	12/23/71/2.0739e-07	12/23/71/2.2742e-07	16/16/80/4.4961e-07	12/23/71/2.3441e-07	12/23/71/2.9048e-07
	0.01	11/21/65/9.2379e-07	16/16/80/2.8202e-07	11/21/65/9.2763e-07	12/23/71/1.4187e-07	16/16/80/2.8219e-07	12/23/71/1.4695e-07	12/23/71/2.0613e-07
21(40)	1	13/25/545/9.2017e-07	18/18/738/2.7625e-07	11/21/461/9.3589e-07	40/125/1725/7.3350e-07	54/79/2239/3.7763e-07	41/121/1761/5.9493e-07	24/65/1025/7.2034e-07
	0.1	12/23/503/8.4921e-07	17/17/697/2.7970e-07	11/21/461/2.0369e-07	14/27/587/2.9724e-07	19/24/784/4.3234e-07	28/93/1213/2.0395e-07	26/61/1101/6.4085e-07
	0.01	12/23/503/7.3679e-07	16/16/656/9.1694e-07	11/21/461/1.7123e-07	14/27/587/6.0719e-07	19/24/784/4.8411e-07	43/133/1853/7.1211e-07	26/61/1101/5.8678e-07
22(1000)	1	9/17/9017/2.2067e-07	12/12/12012/4.1637e-07	11/21/11021/6.0991e-07	10/19/10019/3.4824e-07	12/12/12012/5.8368e-07	12/23/12023/1.6934e-07	9/17/9017/3.2438e-07
	10	11/21/11021/6.0840e-07	15/15/15015/8.1423e-07	14/27/14027/1.6853e-07	15/29/15029/9.1611e-07	17/17/17017/4.2302e-07	17/39/17039/6.3083e-07	11/21/11021/5.9859e-07
	100	14/27/14027/9.0056e-08	19/19/19019/1.9873e-07	16/31/16031/3.3134e-07	23/45/23045/8.8539e-08	24/24/24024/1.5810e-07	21/41/21041/4.3271e-07	14/27/14027/8.7301e-08
25(1000)	1	23/45/23045/1.0206e-07	32/32/32032/1.3728e-07	23/45/23045/1.0206e-07	23/45/23045/1.5163e-07	32/32/32032/1.3979e-07	23/45/23045/1.2166e-07	19/37/19037/7.3627e-08
	10	24/47/24047/2.3509e-07	33/33/33033/7.3147e-07	24/47/24047/2.3509e-07	25/49/25049/1.5153e-07	34/34/34034/1.8511e-07	25/49/25049/6.4208e-08	-
	100	27/53/27053/8.8645e-08	37/37/370374.7574e-07	27/53/27053/8.8645e-08	32/63/32063/1.4662e-07	41/41/41041/1.8216e-07	30/59/30059/9.9380e-07	-
26(1000)	0.1	3/5/3005/9.0082e-07	12/27/12027/7.0362e-07	14/45/14045/ 1.0042e-07	3/5/3005/9.0080e-07	13/33/13033/5.7576e-07	13/53/13053/3.6783e-07	18/63/18063/8.5399e-07
	1	9/41/9041/5.5454e-07	24/39/24039/4.4218e-07	20/75/20075/4.2129e-07	8/45/8045/2.7416e-07	23/37/23037/4.1376e-07	12/49/12049/6.9998e-07	10/25/10025/7.3420e-07
	10	15/51/15051/4.0712e-07	27/44/27044/6.3816e-07	23/85/23085/2.1941e-07	21/83/21083/6.3086e-07	32/62/32062/7.6224e-07	21/89/21089/8.8656e-07	14/45/14045/2.7737e-07
	100	15/53//15053/9.6801e-07	-	-	49/169/49169/8.9963e-07	-	-	-
27(1000)	1	13/25/13025/4.8991e-07	17/17/17017/3.7863e-07	13/25/13025/2.7587e-07	47/131/47131/7.0206e-07	53/79/53079/2.2831e-07	25/65/25065/4.3025e-07	25/69/25069/9.0735e-07
	0.1	14/27/14027/2.2496e-07	18/18/18018/3.0361e-07	13/25/13025/5.6833e-07	44/111/44111/4.8945e-07	37/50/37050/2.9880e-07	36/97/36097/2.0433e-07	30/63/30063/5.0488e-07
	0.01	14/27/14027/2.4074e-07	18/18/18018/3.2802e-07	13/25/13025/6.0491e-07	43/107/43107/8.6428e-07	60/89/60089/7.8275e-08	35/119/35119/7.7313e-07	16/31/16031/2.8992e-07
28(1000)	1	4/7/4007/2.8592e-08	3/3/3003/1.7819e-07	3/5/3005/1.5575e-07	4/7/4007/2.0251e-09	3/3/3003/1.1654e-09	3/5/3005/1.8553e-10	5/9/5009/1.8354e-07
	10	5/9/5009/6.2302e-07	5/5/5005/1.4688e-07	5/9/5009/1.2828e-07	4/7/4007/6.2984e-07	4/4/4004/1.6141e-08	4/7/4007/3.2204e-09	8/15/8015/3.9845e-07
	100	11/21/11021/4.5519e-08	12/12/12012/4.1989e-08	10/19/10019/6.4549e-07	10/19/10019/5.7089e-07	11/11/11011/1.0689e-07	9/17/9017/3.6333e-07	14/27/14027/3.4697e-07
30(1000)	1	83/293/83293/8.0125e-07	98/193/98193/9.4315e-07	99/379/99379/8.2873e-07	128/503/128503/8.3373e-07	128/246/128246/8.8849e-07	94/349/94349/8.7405e-07	97/341/97341/6.8246e-07
	0.1	12/37/12037/7.7718e-07	82/158/82158/9.5928e-07	92/347/92347/8.3789e-07	119/467/119467/8.3257e-07	107/205/107205/8.9413e-07	87/317/87317/8.8543e-07	52/165/52165/9.5061e-07

From Tables 2 and 3, we see that the nonmonotone LM method with $\delta = 1$ always outperforms the nonmonotone LM method with $\delta = 2$ in terms of the number of iterations and the total number of calculations. We observe from Tables 2 and 3 that our proposed nonmonotone LM method with correction (NLMC) is superior to the LMC method [19] and MTLM method [3] since it requires fewer numbers of Jacobian calculations and function calculations for almost all test problems. From the results showed in Tables 2 and 3, we find that the nonmonotone LM method with correction (NLMC) achieves nearly the same numerical results as AMLM method [14], and in some cases even better. In a word, our proposed nonmonotone LM method is efficient in both the first set problems with $\mathrm{rank}(F'(x)) = n-1$ and the second set problems with $\mathrm{rank}(F'(x)) = n-2$. This means that our new choice of the nonmonotone LM parameter is satisfactory for the singular nonlinear equations. Hence our new Algorithm 2.1 is promising and efficient.

Table 3: Numerical results on the second singular test set with $\mathrm{rank}(F'(x^*)) = n - 2$.

Label(n)	x_0	NLMC1	LMC1[19]	AMLM1[14]	NLMC2	LMC2 [19]	AMLM2[14]	MTLM [3]
		NJ/NF/NT/F	NJ/NF/NT/F	NJ/NF/NT/F	NJ/NF/NT/F	NJ/NF/NT/F	NJ/NF/NT/F	NJ/NF/NT/F
2(2)	1	8/15/31/3.6049e-07	11/11/33/3.0203e-07	8/15/31/3.6049e-07	8/15/31/3.6101e-07	11/11/33/3.0203e-07	8/15/31/3.6101e-07	8/15/31/3.6276e-07
	10	13/25/51/9.2140e-08	17/17/51/6.0757e-07	13/25/51/9.2140e-08	13/25/51/1.5818e-07	17/17/51/6.1883e-07	13/25/51/1.6554e-07	13/25/51/9.2137e-08
	100	17/33/67/1.3161e-07	23/23/69/3.6215e-07	17/33/67/1.3161e-07	20/39/79/2.4130e-07	24/24/72/9.6027e-07	20/39/79/5.9042e-07	17/33/67/1.3161e-07
7(3)	1	11/21/54/7.2790e-07	18/22/76/8.3518e-07	13/25/64/5.0159e-07	10/19/49/8.2223e-07	20/26/86/5.1111e-07	22/57/123/6.0161e-07	13/25/64/7.5784e-07
	10	14/27/69/2.4337e-07	23/32/101/4.9805e-07	14/27/69/2.7060e-07	24/67/139/3.3247e-07	22/22/88/6.3583e-07	15/29/74/5.7237e-07	13/25/64/4.8473e-07
	100	15/29/74/4.4193e-07	19/19/76/5.7073e-07	15/29/74/4.6010e-07	33/101/200/3.6856e-07	26/26/104/5.7127e-07	25/49/124/5.1838e-07	13/25/64/4.1598e-07
14(4)	1	11/21/65/7.6509e-07	-	11/21/65/6.6710e-07	-	-	11/31/75/6.8704e-07	11/21/65/7.6590e-07
	0.1	10/19/59/5.7468e-07	391/771/2335/9.6881e-07	10/19/59/5.0107e-07	-	-	12/35/83/4.7693e-07	10/19/59/5.7761e-07
	0.01	10/19/59/3.6077e-07	-	10/19/59/3.1455e-07	316/1261/2525/9.8195e-07	-	12/23/71/5.3315e-07	10/19/59/3.6338e-07
21(40)	1	9/17/377/5.7002e-07	13/13/533/1.3859e-07	9/17/377/5.7002e-07	9/17/377/6.4810e-07	13/13/533/1.3882e-07	9/17/377/6.4665e-07	9/17/377/5.7285e-07
	0.1	9/17/377/7.5210e-08	12/12/492/1.4629e-07	9/17/377/7.5210e-08	9/17/377/7.7834e-08	12/12/492/1.4631e-07	9/17/377/7.7822e-08	9/17/377/7.6285e-08
	0.01	9/17/377/5.5484e-08	11/11/451/8.6335e-07	9/17/377/5.5484e-08	9/17/377/5.7056e-08	11/11/451/8.6342e-07	9/17/377/5.7050e-08	9/17/377/5.6360e-08
22(1000)	1	9/17/9017/2.2067e-07	12/12/12012/4.1637e-07	9/17/9017/2.1668e-07	10/19/10019/3.4824e-07	12/12/12012/5.8368e-07	10/19/10019/1.5414e-07	9/17/9017/3.2438e-07
	10	11/21/11021/6.0840e-07	15/15/15015/8.1423e-07	11/21/11021/5.8654e-07	15/29/15029/9.1611e-07	17/17/17017/4.2302e-07	16/31/16031/1.9828e-07	11/21/11021/5.9859e-07
	100	14/27/14027/9.0056e-08	19/19/19019/1.9873e-07	14/27/14027/7.5916e-08	23/45/23045/8.8539e-08	24/24/24024/1.5810e-07	23/45/23045/2.3026e-07	14/27/14027/8.7301e-08
25(1000)	1	23/45/23045/1.0458e-07	32/32/32032/1.3728e-07	23/45/23045/1.0458e-07	23/45/23045/1.5832e-07	32/32/32032/1.3975e-07	23/45/23045/1.5844e-07	-
	10	24/47/24047/2.3407e-07	33/33/33033/7.2748e-07	24/47/24047/2.3534e-07	25/49/25049/1.5231e-07	34/34/34034/1.8587e-07	25/49/25049/2.4891e-07	-
	100	27/53/27053/1.3302e-07	-	27/53/27053/9.4614e-08	32/63/32063/1.2762e-07	-	32/63/32063/2.9222e-07	-
26(1000)	0.1	3/5/3005/9.0082e-07	11/27/11027/4.7312e-07	15/47/15047/4.8865e-07	3/5/3005/9.0080e-07	12/30/12030/1.2340e-07	17/55/17055/1.7775e-07	22/71/22071/8.7887e-07
	1	10/41/10041/6.5691e-07	25/41/25041/5.6064e-07	10/33/10033/9.4869e-07	9/47/9047/8.7609e-07	22/35/22035/6.5942e-07	19/57/19057/5.5136e-07	10/25/10025/7.3416e-07
	10	15/47/15047/9.1110e-07	32/46/32046/9.8924e-07	17/55/17055/1.5212e-07	36/123/36123/2.3445e-07	38/64/38064/8.5258e-07	24/89/24089/4.3313e-07	15/45/15045/1.4835e-07
	100	15/55/15055/4.6043e-07	370/564/370564/7.1329e-07	-	18/77/18077/8.0058e-07	-	-	-
27(1000)	1	8/15/8015/6.3638e-07	11/11/11011/5.2063e-07	8/15/8015/6.3638e-07	8/15/8015/7.4693e-07	11/11/11011/5.2207e-07	8/15/8015/7.4421e-07	8/15/8015/6.8220e-07
	0.1	9/17/9017/2.3018e-07	12/12/12012/4.4637e-07	9/17/9017/2.3018e-07	9/17/9017/3.7943e-07	12/12/12012/4.6035e-07	9/17/9017/3.9914e-07	9/17/9017/2.3623e-07
	0.01	9/17/9017/2.6447e-07	12/12/12012/5.1286e-07	9/17/9017/2.6446e-07	9/17/9017/4.5327e-07	12/12/12012/5.3184e-07	9/17/9017/4.8504e-07	9/17/9017/2.7094e-07
28(1000)	1	4/7/4007/2.8582e-08	3/3/3003/1.7819e-07	4/7/4007/4.7301e-08	4/7/4007/1.7439e-07	3/3/3003/2.9206e-08	4/7/4007/6.7207e-08	5/9/5009/1.8353e-07
	10	5/9/5009/6.2302e-07	5/5/5005/1.4688e-07	6/11/6011/7.6640e-08	4/7/4007/6.2983e-07	4/4/4004/1.9436e-07	5/9/5009/1.0576e-07	8/15/8015/3.9844e-07
	100	11/21/11021/4.1498e-07	12/12/12012/3.3942e-07	11/21/11021/9.2516e-07	11/21/11021/1.0070e-08	11/11/11011/6.2666e-07	11/21/11021/6.3608e-08	15/29/15029/3.8064e-08
30(1000)	1	83/293/83293/8.0151e-07	98/193/98193/9.4295e-07	12/45/12045/5.9457e-07	128/503/128503/8.3122e-07	128/246/128246/8.8615e-07	14/59/14059/6.0027e-07	97/341/97341/6.8227e-07
	0.1	12/37/12037/7.7807e-07	82/158/82158/9.5924e-07	103/393/103393/8.6624e-07	119/467/119467/8.3080e-07	107/205/107205/8.9318e-07	61/191/61191/9.6308e-07	52/165/52165/9.5079e-07

6 Conclusions

In this paper, we propose a nonmonotone LM method with correction to solve systems of nonlinear equations for which its global convergence and cubic convergence rate, under mild assumptions, are established. Preliminary numerical results illustrate that the proposed algorithm is efficient and robust.

Acknowledgment

The authors would like to thank the editor and the anonymous referees for their detailed comments which greatly improve the presentation.

References

[1] M. Ahookhosh and K. Amini, A nonmonotone trust region method with adaptive radius for unconstrained optimization, *Comput. Math. Appl.* 60 (2010) 411–422.

[2] M. Ahookhosh and K. Amini, An efficient nonmonotone trust-region method for unconstrained optimization, *Numer. Algorithms* 59 (2012) 523–540.

[3] K. Amini and F. Rostami, A modified two steps Levenberg-Marquardt method for nonlinear equations, *J. Comput. Appl. Math.* 288 (2015) 341–350.

[4] J. Baumeister, B. Kaltenbacher and A. Leitão, On Levenberg-Marquardt-Kaczmarz iterative methods for solving systems of nonlinear ill-posed equations, *Inverse Probl. Imag.* 4 (2010) 335–350.

[5] R. Behling and A. Iusem, The effect of calmness on the solution set of systems of nonlinear equations, *Math. Program.* 137 (2011) 155–165.

[6] L. Chen, A higher-order modified Levenberg-Marquardt method for systems of nonlinear equations with fourth-order convergence, *Appl. Math. Comput.* 285 (2016) 79–93.

[7] L. Chen and Y.F. Ma, Shamanskii-like Levenberg-Marquardt method with a new line search for systems of nonlinear equations, *J. Syst. Sci. Complex.* 33 (2020) 1694–1707.

[8] W.L. Cruz, J.M. Martínez and M. Raydan, Spectral residual method without gradient information for solving large-scale nonlinear systems of equations, *Math. Comput.* 75 (2006) 1449–1466.

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

[10] E.D. Dolan and J.J.Moŕe, Benchmarking optimization software with performance profiles, *Math. Program.* 91 (2002) 201–213.

[11] H. Esmaeili and M. Kimiaei, A new adaptive trust-region method for system of nonlinear equations, *Appl. Math. Model.* 19 (2014) 469–490.

[12] H. Esmaeili and M. Kimiaei, An effective adaptive trust-region method for systems of nonlinear equations, *Int. J. Comput. Math.* 92 (2015) 151–166.

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

[14] J.Y. Fan, Accelerating the modified Levenberg-Marquardt method for nonlinear equations, *Math. Comput.* 83 (2014) 1173–1187.

[15] J.Y. Fan, J.C. Huang and J.Y. Pan, An adaptive multi-step Levenberg-Marquardt method, *J. Sci. Comput.* 78 (2019) 531–548.

[16] J.Y. Fan, J.Y. Pan, A note on the Levenberg-Marquardt parameter, *Appl. Math. Comput.* 207 (2009) 351–359.

[17] J.Y. Fan and J.Y, Pan, Convergence properties of a self-adaptive Levenberg-Marquardt algorithm under local error bound condition, *Comput. Optim. Appl.* 34 (2006) 47–62.

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

[19] J.Y. Fan and J.L. Zeng, A Levenberg-Marquardt algorithm with correction for singular system of nonlinear equations, *Appl. Math. Comput.* 219 (2013) 9438–9446.

[20] Z.W. Fu, B. Han and Y. Chen, Levenberg-Marquardt method with general convex penalty for nonlinear inverse problems, *J. Comput. Appl. Math.* 404 (2022): 113771.

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

[22] L. Grippo, F. Lampariello and S. Lucidi, A truncated Newton method with nonmonotone line search for unconstrained optimizations, *J. Optim. Theory. Appl.* 60 (1989) 401–419.

[23] B.H. Huang and C.F. Ma, A Shamanskii-like self-adaptive Levenberg-Marquardt method for nonlinear equations, *Comput. Math. Appl.* 77 (2019) 357–373.

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

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

[26] L. Lukšan and J. Vlček, Sparse and partially separable test problems for unconstrained and equality constrained optimization, Techical Report, No. 767, January 1999.

[27] C.F. Ma and L.H. Jiang, Some research on Levenberg-Marquardt method for the nonlinear equations, *Appl. Math. Comput.* 184 (2007) 1032–1040.

[28] J.J. Moŕe, B.S. Garbow and K.H. Hillstrom, Testing unconstrained optimization software, *ACM Trans. Math. Software* 7 (1981) 17–41.

[29] M.J.D. Powell, Convergence properties of a class of minimization algorithms, in: *Nonlinear Programming 2*, O.L. Mangasarian, R.R. Meyer, S.M. Robinson (eds.), Academic Press, New York, 1975, pp. 1–27.

[30] R.B. Schnabel and P.D. Frank, Tensor methods for nonlinear equations, *SIAM J. Numer. Anal.* 21 (1984) 815–843.

[31] V.E. Shamanskii, A modification of Newton's method, *Ukr. Mat. Zh.* 19 (1967) 133–138.

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

[33] Z.J. Wang, L. Cai, Y.F. Su and Z. Peng, An inexact affine scaling Levenberg-Marquardt method under local error bound conditions, *Acta Math. Appl. Sinica* 35 (2019) 830–844.

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

[35] X. Yang, A higher-order Levenberg-Marquardt method for nonlinear equations, *Appl. Math. Comput.* 219 (2013) 10682–10694.

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

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

Manuscript received 2 September 2022
revised 10 August 2023
accepted for publication 15 August 2023

Baohua Huang
School of Mathematics and Statistics & Key Laboratory of
Analytical Mathematics and Applications (Ministry of Education)
Fujian Normal University, Fuzhou 350117, P.R. China
E-mail address: baohuahuang@fjnu.edu.cn

Changfeng Ma
School of Mathematics and Statistics
Fujian Normal University
Fuzhou 350117, P.R. China
E-mail address: macf@fjnu.edu.cn