AIMS Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 9(9): 24610–24635.
DOI:10.3934/math.20241199
Received: 11 June 2024
Revised: 11 August 2024
Accepted: 15 August 2024
Published: 22 August 2024

Research article

An accelerated adaptive two-step Levenberg–Marquardt method with the modified Metropolis criterion

Dingyu Zhu, Yueting Yang* and Mingyuan Cao*

School of Mathematics and Statistics, Beihua University, Jilin 132013, China

* **Correspondence:** Email: yyt2858@163.com, cmy0918@beihua.edu.cn.

Abstract: In this paper, aiming at the nonlinear equations, a new two-step Levenberg–Marquardt method was proposed. We presented a new Levenberg–Marquardt parameter to obtain the trial step. A new modified Metropolis criterion was used to adjust the upper bound of the approximate step. The convergence of the method was analyzed under the Hölderian local error bound condition and the Hölderian continuity of the Jacobian. Numerical experiments showed that the new algorithm is effective and competitive in the numbers of functions, Jacobian evaluations and iterations.

Keywords: nonlinear equations; Levenberg–Marquardt method; Metropolis criterion; Hölderian local error bound; Hölderian continuity
Mathematics Subject Classification: 65K05, 90C30

1. Introduction

The nonlinear equation is a popular topic in many research fields [1–5], including engineering design, physics, computational science, etc. However, with the increase in data scale and problem complexity, solving nonlinear equations has become incrementally challenging. Therefore, studying effective numerical methods to solve nonlinear equations has highly theoretical and practical significance.

We consider solving nonlinear equations

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

where $F(x) : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable and the solution set of (1.1) is nonempty denoted by X^*. There are many numerical methods [6–12] to solve nonlinear equations. Among them, the Levenberg–Marquardt (LM) method [13, 14] has attracted much attention by introducing the LM regularizer into the Gauss–Newton method, which enables the algorithm to be well-defined when the Jacobian is singular or close to singular. It computes the LM step $\tilde{d}_k$ as

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

where $F_k = F(x_k)$ and $J_k = F'(x_k)$ is the Jacobian of $F(x)$ at x_k, $\lambda_k > 0$ is an appropriate LM parameter updated with each iteration, and $I \in \mathbb{R}^{n\times n}$ is the identity matrix. Throughout the paper, $\|\cdot\|$ denotes the Euclidean norm.

The choice of the LM parameter is essential for the LM method. Yamashita and Fukushima [15] proved that the LM method had the quadratic convergence rate under the local error bound condition when $\lambda_k = \|F_k\|^2$. Fan and Yuan [16] proposed $\lambda_k = \|F_k\|$, which overcame the shortcoming that the LM step was too small when the iteration x_k was far away from the solution. Subsequently, Fan [17] chose λ_k as $\mu_k\|F_k\|$, in which μ_k was updated by a trust region technique. Amini [18] proposed the LM parameter $\frac{\mu_k\|F_k\|}{1+\|F_k\|}$, and proved the convergence under the local error bound condition. On the other hand, Ma and Jiang [19] chose the LM parameter as $\theta\|F_k\| + (1-\theta)\|J_k^T F_k\|$ with $\theta \in [0, 1]$ and obtained the quadratic convergence rate under the local error bound condition. Fan and Pan [20] proposed the LM parameter

$$\lambda_k = \mu_k \left(\theta\|F_k\| + (1-\theta)\|J_k^T F_k\|\right), \tag{1.3}$$

and preserved the quadratic convergence. From this, we can find that the LM parameter is an important component of algorithm research and deserves further study.

To improve the convergence rate and efficiency of the algorithm, Fan [21] proposed the modified LM algorithm with the LM step $\tilde{d}_k$ in (1.2) and the approximate step

$$\hat{d}_k = -(J_k^T J_k + \lambda_k I)^{-1} J_k^T F(y_k), \tag{1.4}$$

where $y_k = x_k + \tilde{d}_k$ and $\lambda_k = \mu_k\|F_k\|^\delta$ with $\delta \in [1, 2]$. Using J_k instead of $J(y_k)$ could effectively save the calculations of the Jacobian. Under the local error bound condition, the modified LM method achieved a cubic convergence. Fan and Zeng [22] introduced a new correction step:

$$\hat{d}_k = (J_k^T J_k + \lambda_k I)^{-1} \lambda_k \tilde{d}_k,$$

where $\lambda_k = \mu_k\|F_k\|^\delta$ with $\delta \in (0, 2]$ and the convergence rate was $\min\{2, 1 + 2\delta\}$ under the same conditions. Above all, the trial step of each iteration became

$$\bar{s}_k = \tilde{d}_k + \hat{d}_k,$$

and the step size was a unit. Then, Fan [23] proposed the accelerated modified LM method, which introduced a line search along $\hat{d}_k$ of (1.4). The step size was the solution of

$$\max_{\alpha\in[1,\hat{\alpha}]} \|F(y_k)\|^2 - \|F(y_k) + \alpha J_k \hat{d}_k\|^2 := \phi(\alpha), \text{ where } \hat{\alpha} > 1. \tag{1.5}$$

By a simple derivation,

$$\tilde{\alpha}_k := \operatorname{argmax} \phi(\alpha) = 1 + \frac{\lambda_k \hat{d}_k^T \hat{d}_k}{\hat{d}_k^T J_k^T J_k \hat{d}_k} > 1, \text{ when } J_k\hat{d}_k \neq \mathbf{0}. \tag{1.6}$$

If $J_k\hat{d}_k$ was close to $\mathbf{0}$, $\tilde{\alpha}_k$ would be too large. An upper bound $\hat{\alpha} > 1$ for α in (1.5) was set and the step size was chosen as $\alpha_k = \min(\tilde{\alpha}_k, \hat{\alpha})$. Moreover, the trust region ratio was introduced by

$$r_k = \frac{Ared_k}{Pred_k} = \frac{\|F_k\|^2 - \|F(x_k + \tilde{d}_k + \alpha_k\hat{d}_k)\|^2}{\|F_k\|^2 - \|F_k + J_k\tilde{d}_k\|^2 + \|F(y_k)\|^2 - \|F(y_k) + \alpha_k J_k\hat{d}_k\|^2}, \tag{1.7}$$

which was used to decide whether to accept the trial step and updated the parameter μ_k. However, the choice strategy of $\hat{\alpha}$ and its influence to the convergence of the algorithm is not mentioned. This inspires us to consider an adaptive updated strategy to the upper bound of the step size in each iteration, which enables the algorithm to preserve the cubic convergence and not increase the computational cost of the Jacobian evaluations. Note that the different choice strategy of λ_k also leads to the different LM method. We will propose a new LM parameter and construct a new two-step LM method with adaptive step size.

When proving the convergence rate, some problems do not satisfy the local error bound condition, but practically satisfy the Hölderian error bound condition. Zhu et al. [24], Wang et al. [25], Zeng et al. [26], and Chen et al. [27] studied the local convergence rate of the LM method under the Hölderian local bound condition with different LM parameters, respectively. To expand the scope and practicality of the algorithm, we devote our research to giving the global and local convergence under the Hölderian conditions.

The aim of our research is to propose an effective accelerated adaptive two-step LM algorithm based on a modified criterion for solving nonlinear equations. The key innovations of this paper are as follows: First, we use the convex combination of $\frac{\|F_k\|}{1+\|F_k\|}$ and $\frac{\|J_k^T F_k\|}{1+\|J_k^T F_k\|}$ as a new LM parameter to update the trial step. Second, considering that different approximate steps may have different upper bounds, we introduce a new modified criterion to update the upper bound of the approximate step size, rather than changing at a constant. Third, the convergence of the new method is proved under the Hölderian local error bound condition and the Hölderian continuity of the Jacobian.

The paper is organized as follows. In next section, a new two-step LM algorithm is described and the global convergence under the Hölderian continuity of the Jacobian is presented. In Section 3, we derive the convergence rate of the new algorithm under the Hölderian local error bound condition and the Hölderian continuity of the Jacobian. In Section 4, numerical experiments show that the new algorithm reduces the numbers of function and Jacobian evaluations. We conclude the paper in Section 5.

2. Algorithm and global convergence

In this section, we propose a novel two-step LM method with a new parameter λ_k. The upper bound of the approximate step size is adjusted by the modified Metropolis criterion. The global convergenceof the new method is proved under the Hölderian continuity of the Jacobian which is weaker than the Lipschitz continuity.

Since the LM step $\tilde{d}_k$ in (1.2) and the approximate step $\hat{d}_k$ in (1.4) rely on the choice of λ_k, we construct a new LM parameter

$$\lambda_k = \mu_k\left(\theta\frac{\|F_k\|}{1+\|F_k\|} + (1-\theta)\frac{\|J_k^T F_k\|}{1+\|J_k^T F_k\|}\right), \text{ where } \theta \in [0,1]. \tag{2.1}$$

When x_k is far from the optimal solution, $\|F_k\|$ and $\|J_k^T F_k\|$ are large enough to make $\frac{\|F_k\|}{1+\|F_k\|}$ and $\frac{\|J_k^T F_k\|}{1+\|J_k^T F_k\|}$ close to 1. At this time, λ_k is close to μ_k. Conversely, when x_k approaches the optimal solution, $\theta\frac{\|F_k\|}{1+\|F_k\|}$ and $(1-\theta)\,\frac{\|J_k^T F_k\|}{1+\|J_k^T F_k\|}$ degenerate into $\theta\|F_k\|$ and $(1-\theta)\,\|J_k^T F_k\|$, which indicates that λ_k is close to the LM parameter mentioned in (1.3). The new LM parameter in (2.1) provides flexibility with the iteration process and enhances the performance of the LM method.

The trial step of the new method is

$$s_k = \tilde{d}_k + \alpha_k \hat{d}_k,$$

where α_k is the step size along $\hat{d}_k$. Unlike the reference [23], we will propose a new upper bound $\hat{\alpha}_k$ of the step size in (1.5). Similar to the Metropolis criterion suggested by [28], we give a new modified Metropolis criterion

$$\bar{\alpha}_k = \begin{cases} 1, & \text{if } |r_{k-1}-1| \le \tau, \\ e^{-\frac{|r_{k-1}-1|}{T_k}}, & \text{otherwise}, \end{cases} \quad \text{with } k \ge 1, \tag{2.2}$$

where $0 < \tau < 1$ represents a sufficiently small constant and T_k is the temperature decreasing to 0 as $k \to \infty$ by the cooling schedule. If $|r_{k-1}-1| \le \tau$, r_{k-1} is close enough to 1, and we set $\bar{\alpha}_k$ as 1. Otherwise, $|r_{k-1}-1| > \tau$, we set $\bar{\alpha}_k = e^{-\frac{|r_{k-1}-1|}{T_k}}$, which can be regarded as a probability and also decreases to 0 as $k \to \infty$. This is similar to the simulated annealing. We define the upper bound of the step size as $\hat{\alpha}_k = 1 + \bar{\alpha}_k$. In each iteration, $\hat{\alpha}_k$ is self-adaptively updated by (2.2). Now, we set the step size along $\hat{d}_k$ as

$$\alpha_k = \min\left(\tilde{\alpha}_k, \hat{\alpha}_k\right), \tag{2.3}$$

where $\tilde{\alpha}_k$ is given by (1.6). Moreover, since $\phi(\alpha)$ has the monotonically increasing property on $[1, \tilde{\alpha}_k]$ and $\alpha_k \in [1, \tilde{\alpha}_k]$, it is easy to find $\phi(\alpha_k) \ge \phi(1)$. This implies

$$\|F(y_k)\|^2 - \|F(y_k) + \alpha_k J_k \hat{d}_k\|^2 \ge \|F(y_k)\|^2 - \|F(y_k) + J_k \hat{d}_k\|^2. \tag{2.4}$$

Based on the above description, we present the accelerated adaptive two-step Levenberg–Marquardt (AATLM) algorithm.

Algorithm 1 AATLM algorithm.

Step 0. Set $x_0 \in \mathbb{R}^n$, $F_0 = F(x_0)$, $J_0 = J(x_0)$, $\varepsilon > 0$, $\mu_0 > m_0 > 0$, $1 \geq \theta \geq 0$, $\bar{\alpha}_0 > 0$, $\tau > 0$, $T_0 = 1$, $C = 0.99$, $1 > q_2 > q_1 > q_0 > 0$, $u > 1$, $a_1 > 1 > a_2 > 0$. Let $k := 0$.

Step 1. If $\|J_k^T F_k\| \leq \varepsilon$, stop, else compute λ_k by (2.1).

Step 2. Solve

$$(J_k^T J_k + \lambda_k I)d = -J_k^T F_k \tag{2.5}$$

to obtain $\tilde{d}_k$, and solve

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

to obtain $\hat{d}_k$. If $\|\hat{d}_k\| \leq \varepsilon$, set $s_k = \tilde{d}_k$, else compute α_k by (1.6), (2.2), (2.3), and set $s_k = \tilde{d}_k + \alpha_k \hat{d}_k$.

Step 3. Compute $r_k = \frac{Ared_k}{Pred_k}$ by (1.7). Set

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{if } r_k \geq q_0, \\ x_k, & \text{otherwise.} \end{cases}$$

Compute F_{k+1} and J_{k+1}.

Step 4. Choose μ_{k+1} as

$$\mu_{k+1} = \begin{cases} a_1 \mu_k, & \text{if } r_k \leq q_1, \\ \mu_k, & \text{if } q_1 < r_k \leq q_2, \\ \max\{a_2 \mu_k, m_0\}, & \text{otherwise.} \end{cases} \tag{2.6}$$

Set $T_{k+1} = CT_k$ and $k := k + 1$, and go to step 1.

Remark 2.1. In Step 2, $\tilde{\alpha}_k$ is computed by (1.6), which is proposed in [23] with $J_k \hat{d}_k \neq \mathbf{0}$. In [23], when $J_k \hat{d}_k$ was close to $\mathbf{0}$, $\hat{\alpha}$ was set as the upper bound of $\tilde{\alpha}_k$. However, the case of $J_k \hat{d}_k = \mathbf{0}$ was not mentioned. Note that, if $\hat{d}_k \neq \mathbf{0}$, then $J_k \hat{d}_k \neq \mathbf{0}$. In fact, if $J_k \hat{d}_k = \mathbf{0}$ holds, from the definition of $\hat{d}_k$, we have

$$-J_k^T F(y_k) = \left(J_k^T J_k + \lambda_k I\right) \hat{d}_k = J_k^T J_k \hat{d}_k + \lambda_k \hat{d}_k = \lambda_k \hat{d}_k \neq \mathbf{0}.$$

Due to $\hat{d}_k$ being the solution of

$$\min_{d \in \mathbb{R}^n} \|F(y_k) + J_k d\|^2 \quad \text{s.t. } \|d\| \leq \Delta_{k,2} := \|\hat{d}_k\|, \tag{2.7}$$

it is easy to obtain

$$\|F(y_k)\|^2 - \|F(y_k) + J_k \hat{d}_k\|^2 \geq \|J_k^T F(y_k)\| \min\left\{\|\hat{d}_k\|, \frac{\|J_k^T F(y_k)\|}{\|J_k^T J_k\|}\right\}.$$

At this time, the left side of the above equation is 0, but the right side is larger than 0. This leads to a contradiction. Therefore, if $\hat{d}_k = \mathbf{0}$, we set $s_k = \tilde{d}_k$, and the algorithm degenerates into a general LM algorithm.

To prove the global convergence of the algorithm, we give the following assumption.

Assumption 2.1. *(a) The Jacobian $J(x)$ is Hölderian continuous of order $\nu \in (0, 1]$, i.e., there exists a positive constant κ_{hj} such that*

$$\|J(y) - J(x)\| \le \kappa_{hj}\|y - x\|^{\nu}, \quad \forall x, y \in \mathbb{R}^n. \tag{2.8}$$

(b) The Jacobian $J(x)$ is bounded above, i.e., there exists a positive constant κ_{bj} such that

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

By using (2.8), we have

$$\begin{aligned}
\|F(y) - F(x) - J(x)(y - x)\| &= \left\| \int_0^1 J(x + t(y - x))(y - x)dt - J(x)(y - x) \right\| \\
&\le \|y - x\| \int_0^1 \|J(x + t(y - x)) - J(x)\|dt \\
&\le \kappa_{hj}\|y - x\|^{1+\nu} \int_0^1 t^{\nu} dt \\
&= \frac{\kappa_{hj}}{1 + \nu}\|y - x\|^{1+\nu}.
\end{aligned} \tag{2.10}$$

Lemma 2.1. *Under the conditions of Assumption 2.1, the sequence $\{x_k\}$ generated by the AATLM algorithm satisfies:*

$$Pred_k \ge \|J_k^T F_k\| \min\left\{\|\tilde{d}_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\} + \|J_k^T F(y_k)\| \min\left\{\|\hat{d}_k\|, \frac{\|J_k^T F(y_k)\|}{\|J_k^T J_k\|}\right\}$$

for all k.

Proof. Since $\tilde{d}_k$ is the solution of the following trust region subproblem,

$$\min_{d \in \mathbb{R}^n} \|F_k + J_k d\|^2 \quad \text{s.t. } \|d\| \le \Delta_{k,1} := \|\tilde{d}_k\|,$$

for any $\beta \in [0, 1]$, it follows:

$$\begin{aligned}
\|F_k\|^2 - \|F_k + J_k \tilde{d}_k\|^2 &\ge \|F_k\|^2 - \left\| F_k - J_k \frac{\beta \Delta_{k,1}}{\|J_k^T F_k\|} J_k^T F_k \right\|^2 \\
&\ge 2\beta \Delta_{k,1} \|J_k^T F_k\| - \beta^2 \Delta_{k,1}^2 \|J_k^T J_k\|.
\end{aligned}$$

Then,

$$\begin{aligned}
\|F_k\|^2 - \|F_k + J_k \tilde{d}_k\|^2 &\ge \max_{0 \le \beta \le 1} \left\{ 2\beta \Delta_{k,1} \|J_k^T F_k\| - \beta^2 \Delta_{k,1}^2 \|J_k^T J_k\| \right\} \\
&\ge \|J_k^T F_k\| \min\left\{\|\tilde{d}_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\}.
\end{aligned} \tag{2.11}$$

If $\hat{d}_k = \mathbf{0}$, (2.11) implies that the conclusion of Lemma 2.1 holds. Otherwise, $\hat{d}_k$ is the solution of (2.7), and it holds that

$$\|F(y_k)\|^2 - \|F(y_k) + J_k \hat{d}_k\|^2 \ge \|F(y_k)\|^2 - \left\| F(y_k) - J_k \frac{\beta \Delta_{k,2}}{\|J_k^T F(y_k)\|} J_k^T F(y_k) \right\|^2$$

$$\geq \quad 2\beta\Delta_{k,2}\|J_k^T F(y_k)\| - \beta^2\Delta_{k,2}^2\|J_k^T J_k\|.$$

According to (2.4), we have

$$\begin{aligned} \|F(y_k)\|^2 - \|F(y_k) + \alpha_k J_k \hat{d}_k\|^2 &\geq \max_{0\leq\beta\leq 1} \left\{2\beta\Delta_{k,2}\|J_k^T F(y_k)\| - \beta^2\Delta_{k,2}^2\|J_k^T J_k\|\right\} \\ &\geq \|J_k^T F(y_k)\|\min\left\{\|\hat{d}_k\|, \frac{\|J_k^T F(y_k)\|}{\|J_k^T J_k\|}\right\}. \end{aligned} \tag{2.12}$$

The conclusion follows from adding (2.11) and (2.12). □

Now, we give the global convergence of the AATLM algorithm.

Theorem 2.1. *Under the conditions of Assumption 2.1, the sequence $\{x_k\}$ generated by the AATLM algorithm satisfies*

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

Proof. We prove by contradiction. Suppose (2.13) is not true. There exist a positive constant δ and infinitely many k such that

$$\|J_k^T F_k\| \geq \delta, \quad \forall k. \tag{2.14}$$

Let the sets of the indices S_1 and S_2 be

$$S_1 = \left\{k \| \|J_k^T F_k\| \geq \delta\right\},$$

$$S_2 = \left\{k \| \|J_k^T F_k\| \geq \frac{\delta}{2} \text{ and } x_{k+1} \neq x_k\right\},$$

where S_1 is an infinite set. Consider the following two cases.
Case 1: S_2 is finite. We have

$$S_3 = \left\{k \| \|J_k^T F_k\| \geq \delta \ \text{ and } \ x_{k+1} \neq x_k\right\}$$

is also finite. Let $\tilde{k}$ be the largest index of S_3, which means $x_{k+1} = x_k$ holds for all $k \in \{k > \tilde{k} | k \in S_1\}$. Define the indicator set

$$S_4 = \left\{k > \tilde{k} \| \|J_k^T F_k\| \geq \delta \ \text{ and } \ x_{k+1} = x_k\right\}.$$

We notice that $\|J_{k+1}^T F_{k+1}\| \geq \delta$ and $x_{k+2} = x_{k+1}$ for all $k \in S_4$. Otherwise, if $x_{k+2} \neq x_{k+1}$, then $k+1 \in S_3$, which means that $\tilde{k}$ is not the largest index of S_3. It is easy to get $k+1 \in S_4$. By induction, $\|J_k^T F_k\| \geq \delta$ and $x_{k+1} = x_k$ hold for all $k > \tilde{k}$.

According to Step 3 in the AATLM algorithm, $r_k < q_0$ means that $x_{k+1} = x_k$ holds for all $k > \tilde{k}$, and from (2.1), (2.5), and (2.6), we obtain:

$$\mu_k \to +\infty \ \text{ and } \ \lambda_k \to +\infty, \tag{2.15}$$

which implies that

$$\tilde{d}_k \to 0.$$

From (2.9), (2.10), (2.7), (2.15), and the definition of $\hat{d}_k$, we find

$$\begin{aligned}
\|\hat{d}_k\| &= \|-(J_k^TJ_k+\lambda_kI)^{-1}J_k^TF(y_k)\| \\
&\le \|(J_k^TJ_k+\lambda_kI)^{-1}J_k^TF_k\| + \|(J_k^TJ_k+\lambda_kI)^{-1}J_k^TJ_k\tilde{d}_k\| + \frac{\kappa_{hj}}{1+\nu}\|\tilde{d}_k\|^{1+\nu}\|(J_k^TJ_k+\lambda_kI)^{-1}J_k^T\| \\
&\le \|\tilde{d}_k\| + \|\tilde{d}_k\| + \frac{\kappa_{hj}\kappa_{bj}}{(1+\nu)\lambda_k}\|\tilde{d}_k\|^{1+\nu} \\
&\le \bar{c}\|\tilde{d}_k\| \qquad (2.16)
\end{aligned}$$

for all sufficiently large k, where $\bar{c}$ is a positive constant. So, we conclude

$$\|s_k\| = \|\tilde{d}_k + \alpha_k\hat{d}_k\| \le (1+\bar{c}\alpha_k)\|\tilde{d}_k\|. \qquad (2.17)$$

On the other hand, it is clear from (2.10) that

$$\begin{cases} \big|\|F(y_k)\| - \|F_k + J_k\tilde{d}_k\|\big| \le \frac{\kappa_{hj}}{1+\nu}\|\tilde{d}_k\|^{1+\nu}, \\ \big|\|F(x_k+s_k)\| - \|F(y_k)+\alpha_kJ_k\hat{d}_k\|\big| \le \frac{\kappa_{hj}}{1+\nu}\|s_k\|^{1+\nu} + \frac{\kappa_{hj}}{1+\nu}\|\tilde{d}_k\|^{1+\nu}, \end{cases}$$

and

$$\begin{cases} \big|\|F(y_k)\| + \|F_k + J_k\tilde{d}_k\|\big| \le 2\|F_k + J_k\tilde{d}_k\| + \frac{\kappa_{hj}}{1+\nu}\|\tilde{d}_k\|^{1+\nu}, \\ \big|\|F(x_k+s_k)\| + \|F(y_k)+\alpha_kJ_k\hat{d}_k\|\big| \le 2\|F_k + J_ks_k\| + \frac{\kappa_{hj}}{1+\nu}\|s_k\|^{1+\nu} + \frac{\kappa_{hj}}{1+\nu}\|\tilde{d}_k\|^{1+\nu}. \end{cases}$$

From the above formulas and Lemma 2.1, (2.10), (2.14), and (2.17), we have

$$\begin{aligned}
|r_k - 1| &= \left|\frac{Ared_k - Pred_k}{Pred_k}\right| \\
&\le \left|\frac{\|F(x_k+s_k)\|^2 - \|F_k+J_k\tilde{d}_k\|^2 + \|F(y_k)\|^2 - \|F(y_k)+\alpha_kJ_k\hat{d}_k\|^2}{\|J_k^TF_k\|\min\left\{\|\tilde{d}_k\|, \frac{\|J_k^TF_k\|}{\|J_k^TJ_k\|}\right\} + \|J_k^TF(y_k)\|\min\left\{\|\hat{d}_k\|, \frac{\|J_k^TF(y_k)\|}{\|J_k^TJ_k\|}\right\}}\right| \\
&\le \frac{\|F_k+J_k\tilde{d}_k\|O\left(\|\tilde{d}_k\|^{1+\nu}\right) + \|F_k+J_ks_k\|O\left(\|\tilde{d}_k\|^{1+\nu}+\|s_k\|^{1+\nu}\right)}{\|\tilde{d}_k\|} \\
&\quad + \frac{O\left(\|\tilde{d}_k\|^{2+2\nu} + \|\tilde{d}_k\|^{1+\nu}\|s_k\|^{1+\nu} + \|s_k\|^{2+2\nu}\right)}{\|\tilde{d}_k\|} \to 0, \qquad (2.18)
\end{aligned}$$

which means that $r_k \to 1$. According to the updating rule of μ_k, we know that there exists a positive constant $M > m_0$, such that $\mu_k < M$ holds for all sufficiently large k, which contradicts with (2.15). Now, we point out that the assumption (2.14) is not true.

Case 2: S_2 is infinite. From Lemma 2.1, (2.10), and the fact that s_k is accepted by the AATLM algorithm, we have

$$\begin{aligned}
\|F_1\|^2 &\ge \sum_{k\in S_2}\left(\|F_k\|^2 - \|F_{k+1}\|^2\right) \ge \sum_{k\in S_2} q_0 Pred_k \\
&\ge \sum_{k\in S_2} q_0\left\{\|J_k^TF_k\|\min\left\{\|\tilde{d}_k\|, \frac{\|J_k^TF_k\|}{\|J_k^TJ_k\|}\right\} + q_0\|J_k^TF(y_k)\|\min\left\{\|\hat{d}_k\|, \frac{\|J_k^TF(y_k)\|}{\|J_k^TJ_k\|}\right\}\right\}
\end{aligned}$$

$$\geq \sum_{k\in S_2} \frac{q_0\delta}{2}\min\left\{\|\tilde{d}_k\|, \frac{\delta}{2\kappa_{bj}^2}\right\}, \tag{2.19}$$

and $x_{k+1} - x_k = 0$ if $k \notin S_2$, which implies that

$$\tilde{d}_k \to 0, \quad k \in S_2, \tag{2.20}$$

and from the definition of $\tilde{d}_k$, we obtain:

$$\lambda_k \to +\infty, \quad k \in S_2. \tag{2.21}$$

Similarly to (2.16) and (2.17), there exists a constant $\tilde{c} > 0$, which makes it true for all sufficiently large $k \in S_2$, so,

$$\|s_k\| = \|\tilde{d}_k + \alpha_k\hat{d}_k\| \leq (1 + \tilde{c}\alpha_k)\,\|\tilde{d}_k\|. \tag{2.22}$$

It follows from (2.19) that

$$\sum_{k\in S_2}\|s_k\| = \sum_{k\in S_2}\|\tilde{d}_k + \alpha_k\hat{d}_k\| < +\infty.$$

Moreover, combining with Assumption 2.1, we get

$$\sum_{k\in S_2}\left|\|J_k^TF_k\| - \|J_{k+1}^TF_{k+1}\|\right| < +\infty.$$

Since (2.14) holds for sufficiently large k, there exists a large $\hat{k}$, such that$\|J_{\hat{k}}^TF_{\hat{k}}\| \geq \delta$, and

$$\sum_{k\in S_2, k\geq\hat{k}}\left|\|J_k^TF_k\| - \|J_{k+1}^TF_{k+1}\|\right| < \frac{\delta}{2}.$$

By induction, we find that $\|J_k^TF_k\| \geq \frac{\delta}{2}$ holds for all $k \geq \hat{k}$, and then, we can derive from (2.19)–(2.22) that

$$\lim_{k\to\infty}\tilde{d}_k = 0 \text{ and } \lim_{k\to\infty}\hat{d}_k = 0,$$

and thus,

$$\mu_k \to +\infty.$$

Similarly, to the analysis of (2.18), we have

$$r_k \to 1.$$

Therefore, there exists a positive constant $M > m_0$ such that $\mu_k < M$ holds for sufficiently large k, which contradicts (2.14). Above all, we get the conclusion immediately. □

Theorem 2.1 indicates that there is $x^* \in X^*$ such that the sequence $\{x_k\}$ generated by the AATLM algorithm converges to x^*. For the sufficient large k, if x_k lies in a neighborhood of x^*, then x_{k+1} and y_k also lie in the neighborhood.

3. Convergence rate of the AATLM algorithm

In this section, we give the properties of the trial step and the boundary of the LM parameter. In order to establish the convergence rate of the AATLM algorithm under the Hölderian local error bound and Hölderian continuity of the Jacobian, we use the following assumption.

Assumption 3.1. *(a) $F(x)$ provides a Hölderian local error bound of order $\gamma \in (0,1]$ in some neighborhoods of $x^* \in X^*$, i.e., there exist constants $c > 0$ and $0 < b < 1$, such that*

$$c\,\mathrm{dist}(x, X^*) \le \|F(x)\|^{\gamma}, \quad \forall x \in N(x^*, b) = \{x \| \|x - x^*\| \le b\}, \tag{3.1}$$

and when $\gamma = 1$, $F(x)$ provides the local error bound.

(b) The Jacobian $J(x)$ is Hölderian continuous of order $\nu \in (0,1]$, i.e., there exists a constant $\kappa_{hj} > 0$ such that

$$\|J(y) - J(x)\| \le \kappa_{hj}\|y - x\|^{\nu}, \quad \forall x, y \in N(x^*, b). \tag{3.2}$$

From (3.2), we immediately have

$$\|F(y) - F(x) - J(x)(y - x)\| \le \frac{\kappa_{hj}}{1+\nu}\|y - x\|^{1+\nu}, \quad \text{where } \forall x, y \in N\left(x^*, \frac{b}{2}\right), \tag{3.3}$$

and there is a constant $\kappa_{bf} > 0$ such that

$$\|F(y) - F(x)\| \le \kappa_{bf}\|y - x\|, \quad \text{where } \forall x, y \in N\left(x^*, \frac{b}{2}\right). \tag{3.4}$$

Moreover, we denote $\bar{x}_k$ as the closest point to x_k in X^*, i.e., $\mathrm{dist}(x_k, X^*) = \|\bar{x}_k - x_k\|$.

Combining the results given by Behling and Iusem [29], we assume that $\mathrm{rank}(J(\bar{x})) = r$ for all $\bar{x} \in N(x^*, b) \cap X^*$. Suppose the singular value decomposition (SVD) of $J(\bar{x}_k)$ is

$$\bar{J}_k = \bar{U}_k\bar{\Sigma}_k\bar{V}_k^T = \left(\bar{U}_1, \bar{U}_2\right)\begin{pmatrix} \bar{\Sigma}_1 & \\ & 0 \end{pmatrix}\begin{pmatrix} \bar{V}_1^T \\ \bar{V}_2^T \end{pmatrix} = \bar{U}_1\bar{\Sigma}_1\bar{V}_1^T,$$

where $\bar{\Sigma}_1 = \mathrm{diag}(\bar{\sigma}_1, ..., \bar{\sigma}_r)$, and $\bar{\sigma}_1 \ge \bar{\sigma}_2 \ge ... \ge \bar{\sigma}_r > 0$. Thus, we obtain:

$$J_k = U_k\Sigma_kV_k^T = (U_1, U_2, U_3)\begin{pmatrix} \Sigma_1 & & \\ & \Sigma_2 & \\ & & 0 \end{pmatrix}\begin{pmatrix} V_1^T \\ V_2^T \\ V_3^T \end{pmatrix} = U_1\Sigma_1V_1^T + U_2\Sigma_2V_2^T,$$

where $\Sigma_1 = \mathrm{diag}(\sigma_1, ..., \sigma_r)$, $\sigma_1 \ge \sigma_2 \ge ... \ge \sigma_r > 0$, and $\Sigma_2 = \mathrm{diag}(\sigma_{r+1}, ..., \sigma_{r+q})$, $\sigma_r \ge \sigma_{r+1} \ge \sigma_{r+2} \ge ... \ge \sigma_{r+q} > 0$. Following from the theory of matrix perturbation [30], and the Hölderian continuity of J_k, we know

$$\|\mathrm{diag}(\Sigma_1 - \bar{\Sigma}_1, \Sigma_2, 0)\| \le \|J_k - \bar{J}_k\| \le \kappa_{hj}\|\bar{x}_k - x_k\|^{\nu},$$

which yields

$$\|\Sigma_1 - \bar{\Sigma}_1\| \le \kappa_{hj}\|\bar{x}_k - x_k\|^{\nu} \text{ and } \|\Sigma_2\| \le \kappa_{hj}\|\bar{x}_k - x_k\|^{\nu}. \tag{3.5}$$

Lemma 3.1. *Under the conditions of Assumption 3.1, if $x_k, y_k \in N\left(x^*, \frac{b}{4}\right)$, and*

$$\nu > \max\left\{2\left(\frac{1}{\gamma}-1\right), \frac{1}{2\gamma}\right\},$$

there exists a constant $c_1 > 0$ such that

$$\|s_k\| \leq c_1 \mathrm{dist}(x_k, X^*). \tag{3.6}$$

Proof. Since $x_k \in N\left(x^*, \frac{b}{4}\right) = \left\{x \| \|x_k - x^*\| \leq \frac{b}{4}\right\}$, it follows from the definition of $\bar{x}_k$ that

$$\|\bar{x}_k - x^*\| \leq \|\bar{x}_k - x_k\| + \|x_k - x^*\| \leq 2\|x_k - x^*\| \leq \frac{b}{2},$$

which means $\bar{x}_k \in N\left(x^*, \frac{b}{2}\right)$.

From the definition of λ_k, we set $\lambda_{1k} = \frac{\mu_k \theta \|F_k\|}{1+\|F_k\|}$, and $\lambda_{2k} = \frac{\mu_k(1-\theta)\|J_k^T F_k\|}{1+\|J_k^T F_k\|}$. Then, together with (3.1) and $\mu_k > m_0$, we have

$$\lambda_{1k} \geq \begin{cases} \frac{\mu_k \theta}{2}\|F_k\| \geq \frac{m_0\theta}{2} c^{\frac{1}{\gamma}} \|\bar{x}_k - x_k\|^{\frac{1}{\gamma}}, & \text{if } \|F_k\| \leq 1; \\ \frac{\mu_k \theta}{2} \geq \frac{m_0 \theta}{2}, & \text{otherwise.} \end{cases}$$

As we know, $\|F_k\|^2 = F_k^T F_k = F_k^T[F(\bar{x}_k) + J_k(\bar{x}_k - x_k)] + F_k^T H_k$, in which $H_k = F_k - F(\bar{x}_k) - J_k(\bar{x}_k - x_k)$. So, we have $F_k^T J_k(\bar{x}_k - x_k) = \|F_k\|^2 - F_k^T H_k$. From the Assumption 3.1, and $\nu > 2\left(\frac{1}{\gamma} - 1\right)$, it is clear that

$$\|J_k^T F_k\| \geq \hat{c}\|\bar{x}_k - x_k\|^{\frac{2}{\gamma}-1}$$

holds for some $\hat{c} > 0$. In the same way, we obtain:

$$\lambda_{2k} \geq \begin{cases} \frac{\mu_k(1-\theta)}{2}\|J_k^T F_k\| \geq \frac{m_0(1-\theta)}{2} \hat{c}\|\bar{x}_k - x_k\|^{\frac{2}{\gamma}-1}, & \text{if } \|J_k^T F_k\| \leq 1; \\ \frac{\mu_k(1-\theta)}{2} \geq \frac{m_0(1-\theta)}{2}, & \text{otherwise.} \end{cases}$$

Thus, we find that the LM parameter λ_k satisfies:

$$\begin{aligned} \lambda_k &= \mu_k\left(\theta \frac{\|F_k\|}{1+\|F_k\|} + (1-\theta)\frac{\|J_k^T F_k\|}{1+\|J_k^T F_k\|}\right) \\ &\geq \max\left\{\frac{m_0\theta}{2}, \frac{m_0\theta}{2} c^{\frac{1}{\gamma}} \|\bar{x}_k - x_k\|^{\frac{1}{\gamma}}\right\} + \max\left\{\frac{m_0(1-\theta)}{2}, \frac{m_0(1-\theta)}{2}\left(\hat{c}\|\bar{x}_k - x_k\|^{\frac{2}{\gamma}-1}\right)\right\} \\ &\geq \grave{c}\|\bar{x}_k - x_k\|^{\frac{1}{\gamma}}, \end{aligned} \tag{3.7}$$

where $\grave{c} > 0$.

In addition, the equivalence problem of (2.11) is

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

which has the optimal solution $\tilde{d}_k$. Combining with (3.7), we have that

$$\|\tilde{d}_k\|^2 \quad \leq \quad \frac{\varphi_{k,1}(\bar{x}_k - x_k)}{\lambda_k}$$

$$
\begin{aligned}
&= \frac{\|F_k + J_k(\bar{x}_k - x_k)\|^2}{\lambda_k} + \|\bar{x}_k - x_k\|^2 \\
&\leq \frac{k_{hj}^2 \|\bar{x}_k - x_k\|^{2+2\nu}}{\grave{c}(1+\nu)^2 \|\bar{x}_k - x_k\|^{\frac{1}{\gamma}}} + \|\bar{x}_k - x_k\|^2 \\
&\leq c_{1,1} \|\bar{x}_k - x_k\|^{2\min\left\{1, 1+\nu-\frac{1}{2\gamma}\right\}}
\end{aligned}
$$

holds for some $c_{1,1} > 0$, which means that

$$
\|\tilde{d}_k\| \leq c_{1,2} \|\bar{x}_k - x_k\|^{\min\left\{1, 1+\nu-\frac{1}{2\gamma}\right\}} \tag{3.8}
$$

holds for some $c_{1,2} > 0$.

By the definition of $\hat{d}_k$ and (3.3), we obtain

$$
\begin{aligned}
\|\hat{d}_k\| &= \| - (J_k^T J_k + \lambda_k I)^{-1} J_k^T F(y_k)\| \\
&\leq \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T F_k\| + \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T J_k \tilde{d}_k\| \\
&\quad + \frac{\kappa_{hj}}{1+\nu} \|\tilde{d}_k\|^{1+\nu} \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T\| \\
&\leq 2\|\tilde{d}_k\| + \frac{\kappa_{hj}}{1+\nu} \|\tilde{d}_k\|^{1+\nu} \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T\|.
\end{aligned} \tag{3.9}
$$

By using the SVD of J_k, we have

$$
\begin{aligned}
\|(J_k^T J_k + \lambda_k I)^{-1} J_k^T\| &= \left\| (V_1, V_2, V_3) \begin{pmatrix} (\Sigma_1^2 + \lambda_k I)^{-1}\Sigma_1 & & \\ & (\Sigma_2^2 + \lambda_k I)^{-1}\Sigma_2 & \\ & & 0 \end{pmatrix} \begin{pmatrix} U_1^T \\ U_2^T \\ U_3^T \end{pmatrix} \right\| \\
&\leq \left\| \begin{pmatrix} (\Sigma_1^2 + \lambda_k I)^{-1}\Sigma_1 & & \\ & (\Sigma_2^2 + \lambda_k I)^{-1}\Sigma_2 & \\ & & 0 \end{pmatrix} \right\| \\
&\leq \left\| \begin{pmatrix} \Sigma_1^{-1} & \\ & \lambda_k^{-1}\Sigma_2 \end{pmatrix} \right\|.
\end{aligned} \tag{3.10}
$$

Due to the sequence $\{x_k\}$ converging to the nonempty solution set X^*, if $\kappa_{hj}\|\bar{x}_k - x_k\|^{\nu} \leq \frac{\bar{\sigma}_r}{2}$ for any sufficiently large k, from the lower bound of λ_k, we get

$$
\|\Sigma_1^{-1}\| \leq \frac{1}{\bar{\sigma}_r - \kappa_{hj}\|\bar{x}_k - x_k\|^{\nu}} \leq \frac{2}{\bar{\sigma}_r},
$$

and

$$
\|\lambda_k^{-1}\Sigma_2\| \leq \frac{\kappa_{hj}\|\bar{x}_k - x_k\|^{\nu}}{\grave{c}\|\bar{x}_k - x_k\|^{\frac{1}{\gamma}}} = \acute{c}\|\bar{x}_k - x_k\|^{\nu - \frac{1}{\gamma}},
$$

where $\acute{c} > 0$ is a constant. Then, combining with (3.9), (3.10), the lower bound of λ_k, and the range of ν, we can deduce

$$
\|\hat{d}_k\| \leq 2\|\tilde{d}_k\| + \frac{\kappa_{hj}}{1+\nu} \|\tilde{d}_k\|^{1+\nu} \|(J_k^T J_k + \lambda_k I)^{-1} J_k^T\|
$$

$$
\begin{aligned}
&\leq 2\|\tilde{d}_k\| + \frac{2\acute{c}\kappa_{hj}}{\bar{\sigma}_r(1+\nu)}\|\tilde{d}_k\|^{1+\nu}\|\bar{x}_k - x_k\|^{\nu-\frac{1}{\gamma}} \\
&\leq 2\|\tilde{d}_k\| + \frac{2\acute{c}\kappa_{hj}c_{1,2}^{1+\nu}}{\bar{\sigma}_r(1+\nu)}\|\bar{x}_k - x_k\|^{\min\left\{1+2\nu-\frac{1}{\gamma},1+3\nu+\nu^2-\frac{\nu}{2\gamma}-\frac{3}{2\gamma}\right\}} \\
&\leq \check{c}\|\bar{x}_k - x_k\|^{\min\{1,\tau\}},
\end{aligned}
$$

where $\check{c} > 0$ is a constant, and

$$
\tau = \min\left\{1 + \nu - \frac{1}{2\gamma}, 1 + 2\nu - \frac{1}{\gamma}, 1 + 3\nu + \nu^2 - \frac{\nu}{2\gamma} - \frac{3}{2\gamma}\right\}. \tag{3.11}
$$

From assumption $\nu > \max\left\{2\left(\frac{1}{\gamma} - 1\right), \frac{1}{2\gamma}\right\}$ and the condition $\nu, \gamma \in (0, 1]$, we know $\nu > \frac{1}{\gamma} - 1$, and $\gamma \in (\frac{2}{3}, 1]$. It is easy to find $\nu \in (\frac{1}{2}, 1]$. As the exponent γ increases, smaller values on the exponent ν are allowed. We obtain:

$$
\tau_1 - 1 = 1 + \nu - \frac{1}{2\gamma} - 1 = \nu - \frac{1}{2\gamma} > 0,
$$

$$
\tau_2 - 1 = 1 + 2\nu - \frac{1}{\gamma} - 1 = 2\left(\nu - \frac{1}{2\gamma}\right) > 0,
$$

$$
\tau_3 - 1 = 1 + 3\nu + \nu^2 - \frac{\nu}{2\gamma} - \frac{3}{2\gamma} - 1 = 3\left(\nu - \frac{1}{2\gamma}\right) + \nu\left(\nu - \frac{1}{2\gamma}\right) > 0,
$$

which implies

$$
\|\tilde{d}_k\| \leq O\left(\|\bar{x}_k - x_k\|\right), \quad \|\hat{d}_k\| \leq O\left(\|\bar{x}_k - x_k\|\right). \tag{3.12}
$$

Due to the definition of s_k, it is easy to know

$$
\|s_k\| = \|\tilde{d}_k + \alpha_k \hat{d}_k\| \leq O\left(\|\bar{x}_k - x_k\|\right).
$$

The proof is complete. □

Lemma 3.2. *Under the conditions of Assumption 3.1, if* $x_k, y_k \in N\left(x^*, \frac{b}{4}\right)$*, and*

$$
\nu > \max\left\{2\left(\frac{1}{\gamma} - 1\right), \frac{1}{2\gamma}\right\},
$$

there exists a constant $M > m_0$*, such that*

$$
\mu_k \leq M \tag{3.13}
$$

holds for all large k.

Proof. We consider the following two cases.
Case 1: If $\|\bar{x}_k - x_k\| \leq \|\tilde{d}_k\|$, it follows from (3.1), (3.3), and $\nu > 2\left(\frac{1}{\gamma} - 1\right)$ that

$$
\begin{aligned}
\|F_k\| - \|F_k + J_k\tilde{d}_k\| &\geq \|F_k\| - \|F_k + J_k(\bar{x}_k - x_k)\| \\
&\geq c^{\frac{1}{\gamma}}\|\bar{x}_k - x_k\|^{\frac{1}{\gamma}} + O\left(\|\bar{x}_k - x_k\|^{1+\nu}\right)
\end{aligned}
$$

$$\geq \quad c_{2,1}\|\bar{x}_k - x_k\|^{\frac{1}{\gamma}}, \tag{3.14}$$

where $c_{2,1} > 0$ is a constant.

Case 2: If $\|\bar{x}_k - x_k\| > \|\tilde{d}_k\|$, it follows from the second and third inequalities of (3.14), that we have

$$\begin{aligned} \|F_k\| - \|F_k + J_k\tilde{d}_k\| &\geq \|F_k\| - \left\| F_k + \frac{\|\tilde{d}_k\|}{\|\bar{x}_k - x_k\|} J_k(\bar{x}_k - x_k) \right\| \\ &\geq \frac{\|\tilde{d}_k\|}{\|\bar{x}_k - x_k\|} \left(\|F_k\| - \|F_k + J_k(\bar{x}_k - x_k)\|\right) \\ &\geq \frac{\|\tilde{d}_k\|}{\|\bar{x}_k - x_k\|} c_{2,1}\|\bar{x}_k - x_k\|^{\frac{1}{\gamma}} \\ &\geq c_{2,1}\|\tilde{d}_k\|\|\bar{x}_k - x_k\|^{\frac{1}{\gamma}-1}. \end{aligned} \tag{3.15}$$

Using the same analysis as (3.14) and (3.15), we deduce

$$\begin{aligned} \|F(y_k)\| - \|F(y_k) + J_k\hat{d}_k\| &\geq \|F(y_k)\| - \|F(y_k) + J_k(\bar{y}_k - y_k)\| \\ &\geq \|F(y_k)\| - \|F(y_k) + J(y_k)(\bar{y}_k - y_k)\| - \|(J_k - J(y_k))(\bar{y}_k - y_k)\| \\ &\geq c^{\frac{1}{\gamma}}\|\bar{y}_k - y_k\|^{\frac{1}{\gamma}} + O\left(\|\bar{y}_k - y_k\|^{1+\nu}\right) + O\left(\|\tilde{d}_k\|^{\nu}\|\bar{y}_k - y_k\|\right) \\ &\geq c_{2,2}\|\bar{y}_k - y_k\|^{\frac{1}{\gamma}}, \end{aligned} \tag{3.16}$$

where $c_{2,2} > 0$ is a constant with $\|\bar{y}_k - y_k\| \leq \|\hat{d}_k\|$, and

$$\begin{aligned} \|F(y_k)\| - \|F(y_k) + J_k\hat{d}_k\| &\geq \|F(y_k)\| - \left\| F(y_k) + \frac{\|\hat{d}_k\|}{\|\bar{y}_k - y_k\|} J_k(\bar{y}_k - y_k) \right\| \\ &\geq \frac{\|\hat{d}_k\|}{\|\bar{y}_k - y_k\|} \left(\|F(y_k)\| - \|F(y_k) + J_k(\bar{y}_k - y_k)\|\right) \\ &\geq \frac{\|\hat{d}_k\|}{\|\bar{y}_k - y_k\|} c_{2,2}\|\bar{y}_k - y_k\|^{\frac{1}{\gamma}} \\ &\geq c_{2,2}\|\hat{d}_k\|\|\bar{y}_k - y_k\|^{\frac{1}{\gamma}-1} \end{aligned} \tag{3.17}$$

holds for $\|\bar{y}_k - y_k\| > \|\hat{d}_k\|$.

Hence, it follows from (3.14)–(3.17), and the definition of $Pred_k$ that

$$Pred_k \quad \geq \quad \|F_k\|\left(\|F_k\| - \|F_k + J_k\tilde{d}_k\|\right) + \|F(y_k)\|\left(\|F(y_k)\| - \|F(y_k) + J_k\hat{d}_k\|\right) \geq C_k,$$

where

$$\begin{aligned} C_k \quad = \quad & c_{2,1}\|F_k\|\min\left\{\|\bar{x}_k - x_k\|^{\frac{1}{\gamma}}, \|\tilde{d}_k\|\|\bar{x}_k - x_k\|^{\frac{1}{\gamma}-1}\right\} \\ & + c_{2,2}\|F(y_k)\|\min\left\{\|\bar{y}_k - y_k\|^{\frac{1}{\gamma}}, \|\hat{d}_k\|\|\bar{y}_k - y_k\|^{\frac{1}{\gamma}-1}\right\}. \end{aligned}$$

From $(J_k^T F_k)^T \tilde{d}_k < 0$, we can derive that $\|F(y_k)\| < \|F_k\|$. Combining (3.1) and (3.3) with (3.12) yields

$$|r_k - 1| \quad = \quad \left|\frac{Ared_k - Pred_k}{Pred_k}\right|$$

$$
\begin{aligned}
&\leq \left|\frac{\|F(x_k+s_k)\|^2-\|F_k+J_k\tilde{d}_k\|^2+\|F(y_k)\|^2-\|F(y_k)+\alpha_k J_k\hat{d}_k\|^2}{C_k}\right| \\
&\leq \frac{\|F_k+J_k\tilde{d}_k\|O\left(\|\tilde{d}_k\|^{1+\nu}\right)+\|F_k+J_ks_k\|O\left(\|\tilde{d}_k\|^{1+\nu}+\|s_k\|^{1+\nu}\right)}{O\left(\|\tilde{d}_k\|\|\bar{x}_k-x_k\|^{\frac{2}{\gamma}-1}\right)} \\
&\quad+\frac{O\left(\|\tilde{d}_k\|^{2+2\nu}+\|\tilde{d}_k\|^{1+\nu}\|s_k\|^{1+\nu}+\|s_k\|^{2+2\nu}\right)}{O\left(\|\tilde{d}_k\|\|\bar{x}_k-x_k\|^{\frac{2}{\gamma}-1}\right)}.
\end{aligned}
$$

Due to $\nu > \max\left\{2\left(\frac{1}{\gamma}-1\right), \frac{1}{2\gamma}\right\}$, it is clear that $r_k \to 1$. Therefore, we conclude that (3.13) is valid from Step 4 in the AATLM algorithm and Lemma 3.2 is proved. □

Lemma 3.3. *Under the conditions of Assumption 3.1, if* $x_k, y_k \in N\left(x^*, \frac{b}{4}\right)$ *and* $\nu > 2\left(\frac{1}{\gamma}-1\right)$*, we have*

$$
\grave{c}\|\bar{x}_k-x_k\|^{\frac{1}{\gamma}} \leq \lambda_k \leq M\theta\kappa_{bf}\|\bar{x}_k-x_k\|+M(1-\theta)\kappa_{bf}^2\|\bar{x}_k-x_k\|,
$$

where $\grave{c} > 0$ is a constant.

Proof. It follows from (3.7) that

$$
\grave{c}\|\bar{x}_k-x_k\|^{\frac{1}{\gamma}} \leq \lambda_k.
$$

By using Lemma 3.2, (3.2), (3.4), and the definition of λ_k, we conclude

$$
\begin{aligned}
\lambda_k &= \mu_k\left(\theta\frac{\|F_k\|}{1+\|F_k\|}+(1-\theta)\frac{\|J_kF_k\|}{1+\|J_kF_k\|}\right) \\
&\leq \mu_k\theta\|F_k\|+\mu_k(1-\theta)\|J_kF_k\| \\
&\leq M\theta\kappa_{bf}\|\bar{x}_k-x_k\|+M(1-\theta)\kappa_{bf}^2\|\bar{x}_k-x_k\|, \qquad (3.18)
\end{aligned}
$$

which means that λ_k is bounded. Above all, we have the conclusion immediately. □

We use the SVD to calculate the convergence rate of the AATLM algorithm. By the SVD of J_k, we get

$$
\tilde{d}_k = -V_1(\Sigma_1^2+\lambda_kI)^{-1}\Sigma_1U_1^TF_k - V_2(\Sigma_2^2+\lambda_kI)^{-1}\Sigma_2U_2^TF_k, \qquad (3.19)
$$

$$
\hat{d}_k = -V_1(\Sigma_1^2+\lambda_kI)^{-1}\Sigma_1U_1^TF(y_k) - V_2(\Sigma_2^2+\lambda_kI)^{-1}\Sigma_2U_2^TF(y_k), \qquad (3.20)
$$

$$
\begin{aligned}
F_k+J_k\tilde{d}_k &= F_k-U_1\Sigma_1(\Sigma_1^2+\lambda_kI)^{-1}\Sigma_1U_1^TF_k-U_2\Sigma_2(\Sigma_2^2+\lambda_kI)^{-1}\Sigma_2U_2^TF_k \\
&= \lambda_kU_1(\Sigma_1^2+\lambda_kI)^{-1}U_1^TF_k+\lambda_kU_2(\Sigma_2^2+\lambda_kI)^{-1}U_2^TF_k+U_3U_3^TF_k, \qquad (3.21)
\end{aligned}
$$

$$
\begin{aligned}
F(y_k)+J_k\hat{d}_k &= F(y_k)-U_1\Sigma_1(\Sigma_1^2+\lambda_kI)^{-1}\Sigma_1U_1^TF(y_k)-U_2\Sigma_2(\Sigma_2^2+\lambda_kI)^{-1}\Sigma_2U_2^TF(y_k) \\
&= \lambda_kU_1(\Sigma_1^2+\lambda_kI)^{-1}U_1^TF(y_k)+\lambda_kU_2(\Sigma_2^2+\lambda_kI)^{-1}U_2^TF(y_k) \\
&\quad+U_3U_3^TF(y_k). \qquad (3.22)
\end{aligned}
$$

Lemma 3.4. *Under the conditions of Assumption 3.1, if* $x_k, y_k \in N\left(x^*, \frac{b}{4}\right)$*, we have*

(a) $\|U_1U_1^TF_k\| \le \kappa_{bf}\|\bar{x}_k - x_k\|$;
(b) $\|U_2U_2^TF_k\| \le (\frac{\kappa_{hj}}{1+\nu} + \kappa_{hj})\|\bar{x}_k - x_k\|^{1+\nu}$;
(c) $\|U_3U_3^TF_k\| \le \frac{\kappa_{hj}}{1+\nu}\|\bar{x}_k - x_k\|^{1+\nu}$.

Proof. We could obtain (a) directly from (3.4). Let $\bar{s}_k = -J_k^+F_k$, where J_k^+ is the pseudo-inverse of J_k and $\bar{s}_k$ is the least squares solution of $\min\|F_k + J_ks\|$. Then, we obtain (c) from (3.3) that

$$\|U_3U_3^TF_k\| = \|F_k + J_k\bar{s}_k\| \le \|F_k + J_k(\bar{x}_k - x_k)\| \le \frac{\kappa_{hj}}{1+\nu}\|\bar{x}_k - x_k\|^{1+\nu}.$$

Let $\tilde{J}_k = U_1\Sigma_1V_1^T$ and $\tilde{s}_k = -\tilde{J}_k^+F_k$, where $\tilde{J}_k^+$ is the pseudo-inverse of $\tilde{J}_k$ and $\tilde{s}_k$ is the least squares solution of $\min\|F_k + \tilde{J}_ks\|$. Together with (3.4) and (3.5) implies

$$\begin{aligned}\|(U_2U_2^TF_k + U_3U_3^TF_k)\| &= \|F_k + \tilde{J}_k\tilde{s}_k\| \\ &\le \|F_k + \tilde{J}_k(\bar{x}_k - x_k)\| \\ &\le \|F_k + J_k(\bar{x}_k - x_k)\| + \|(\tilde{J}_k - J_k)(\bar{x}_k - x_k)\| \\ &\le \frac{\kappa_{hj}}{1+\nu}\|\bar{x}_k - x_k\|^{1+\nu} + \|(U_2\Sigma_2V_2^T)(\bar{x}_k - x_k)\| \\ &\le \frac{\kappa_{hj}}{1+\nu}\|\bar{x}_k - x_k\|^{1+\nu} + \kappa_{hj}\|\bar{x}_k - x_k\|^{\nu}\|\bar{x}_k - x_k\| \\ &\le (\frac{\kappa_{hj}}{1+\nu} + \kappa_{hj})\|\bar{x}_k - x_k\|^{1+\nu},\end{aligned}$$

which means that we obtain (b) from the orthogonality of U_2 and U_3. The proof is complete. □

Lemma 3.5. *Under the conditions of Assumption 3.1, if* $x_k, y_k \in N\left(x^*, \frac{b}{4}\right)$ *and* $\nu > \max\left\{2\left(\frac{1}{\gamma} - 1\right), \frac{1}{2\gamma}\right\}$, *we have*

(a) $\|U_1U_1^TF(y_k)\| \le O\left(\|\bar{x}_k - x_k\|^{1+\nu}\right)$;
(b) $\|U_2U_2^TF(y_k)\| \le O\left(\|\bar{x}_k - x_k\|^{\nu+\gamma(1+\nu)}\right)$;
(c) $\|U_3U_3^TF(y_k)\| \le O\left(\|\bar{x}_k - x_k\|^{\nu+\gamma(1+\nu)}\right)$.

Proof. From (3.21), Lemmas 3.3 and 3.4, and the range of ν, we have

$$\begin{aligned}\|F_k + J_k\tilde{d}_k\| &\le \lambda_k\|\Sigma_1^2\|^{-1}\|U_1U_1^TF_k\| + \|U_2U_2^TF_k\| + \|U_3U_3^TF_k\| \\ &\le O\left(\|\bar{x}_k - x_k\|^2\right) + O\left(\|\bar{x}_k - x_k\|^{1+\nu}\right) \\ &\le O\left(\|\bar{x}_k - x_k\|^{1+\nu}\right),\end{aligned} \tag{3.23}$$

and from (3.3), (3.8), and (3.23), we have

$$\begin{aligned}\|F(y_k)\| &= \|F(x_k + \tilde{d}_k)\| \\ &\le \|F_k + J_k\tilde{d}_k\| + \frac{\kappa_{hj}}{1+\nu}\|\tilde{d}_k\|^{1+\nu} \\ &\le O\left(\|\bar{x}_k - x_k\|^{1+\nu}\right) + O\left(\|\bar{x}_k - x_k\|^{1+\nu}\right) \\ &= O\left(\|\bar{x}_k - x_k\|^{1+\nu}\right).\end{aligned}$$

Thus, it is clear that

$$\|U_1U_1^TF(y_k)\| \le \|F(y_k)\| \le O\left(\|\bar{x}_k - x_k\|^{1+\nu}\right),$$

which indicates that the following condition of the Hölderian local error bound,

$$\|\bar{y}_k - y_k\| \leq \frac{1}{c}\|F(y_k)\|^\gamma \leq O\left(\|\bar{x}_k - x_k\|^{\gamma(1+\nu)}\right), \tag{3.24}$$

is obtained.

Then, we let $\bar{p}_k = -J_k^+ F(y_k)$, and $\bar{p}_k$ is the least squares solution of $\min\|F(y_k) + J_k p\|$. From (3.2), (3.3), (3.9), (3.24), and the range of ν, we have

$$\begin{aligned}
\|U_3 U_3^T F(y_k)\| &= \|F(y_k) + J_k \bar{p}_k\| \\
&\leq \|F(y_k) + J_k(\bar{y}_k - y_k)\| \\
&\leq \|F(y_k) + J(y_k)(\bar{y}_k - y_k)\| + \|(J_k - J(y_k))(\bar{y}_k - y_k)\| \\
&\leq \frac{\kappa_{hj}}{1+\nu}\|\bar{y}_k - y_k\|^{1+\nu} + \kappa_{hj}\|\tilde{d}_k\|^\nu \|\bar{y}_k - y_k\| \\
&\leq O\left(\|\bar{x}_k - x_k\|^{\min\left\{\gamma(1+\nu)^2, \nu+\gamma(1+\nu)\right\}}\right) \\
&\leq O\left(\|\bar{x}_k - x_k\|^{\nu+\gamma(1+\nu)}\right).
\end{aligned}$$

Let $\tilde{J}_k = U_1 \Sigma_1 V_1^T$ and $\tilde{p}_k = -\tilde{J}_k^+ F(y_k)$, where $\tilde{p}_k$ is the least squares solution of $\min\|F(y_k) + \tilde{J}_k p\|$. It follows from (3.2), (3.3), (3.6), (3.8), (3.24), and the range of ν that

$$\begin{aligned}
\|(U_2 U_2^T + U_3 U_3^T) F(y_k)\| &= \|F(y_k) + \tilde{J}_k p_k\| \\
&\leq \|F(y_k) + \tilde{J}_k(\bar{y}_k - y_k)\| \\
&\leq \|F(y_k) + J(y_k)(\bar{y}_k - y_k)\| + \|(\tilde{J}_k - J(y_k))(\bar{y}_k - y_k)\| \\
&\leq \frac{\kappa_{hj}}{1+\nu}\|\bar{y}_k - y_k\|^{1+\nu} + \|(J_k - J(y_k))(\bar{y}_k - y_k)\| + \|U_2 \Sigma_2 V_2^T(\bar{y}_k - y_k)\| \\
&\leq \frac{\kappa_{hj}}{1+\nu}\|\bar{y}_k - y_k\|^{1+\nu} + \kappa_{hj}\|\tilde{d}_k\|^\nu \|\bar{y}_k - y_k\| + \kappa_{hj}\|\bar{x}_k - x_k\|^\nu \|\bar{y}_k - y_k\| \\
&\leq O\left(\|\bar{x}_k - x_k\|^{\min\left\{\gamma(1+\nu)^2, \nu+\gamma(1+\nu)\right\}}\right) \\
&\leq O\left(\|\bar{x}_k - x_k\|^{\nu+\gamma(1+\nu)}\right),
\end{aligned}$$

and then, together with the orthogonality of U_2 and U_3, we obtain (b) and Lemma 3.5 is proved. □

Theorem 3.1. *Under the conditions of Assumption 3.1, if $x_k, y_k \in N\left(x^*, \frac{b}{4}\right)$, $\nu > 2\left(\frac{1}{\gamma} - 1\right)$, and $\nu > \frac{1}{2\gamma}$, the sequence $\{x_k\}$ generated by the AATLM algorithm converges to the solution set of (1.1) with order $\nu\gamma + \gamma^2(1+\nu)$.*

Proof. From (3.5), (3.20), Lemma 3.5, and the upper bound of $\|\lambda_k^{-1}\Sigma_2\|$, we have

$$\begin{aligned}
\|\hat{d}_k\| &= \| - V_1(\Sigma_1^2 + \lambda_k I)^{-1}\Sigma_1 U_1^T F(y_k) - V_2(\Sigma_2^2 + \lambda_k I)^{-1}\Sigma_2 U_2^T F(y_k)\| \\
&\leq \|\Sigma_1^{-1}\|\|U_1^T F(y_k)\| + \|\lambda_k^{-1}\Sigma_2\|\|U_2^T F(y_k)\| \\
&\leq O\left(\|\bar{x}_k - x_k\|^{1+\nu}\right) + O\left(\|\bar{x}_k - x_k\|^{2\nu+\gamma(1+\nu)-\frac{1}{\gamma}}\right) \\
&\leq O\left(\|\bar{x}_k - x_k\|^{\min\left\{1+\nu, 2\nu+\gamma(1+\nu)-\frac{1}{\gamma}\right\}}\right).
\end{aligned} \tag{3.25}$$

It follows from (3.18), (3.22), and Lemma 3.5 that

$$\|F(y_k) + \alpha_k J_k \hat{d}_k\| \leq \|F(y_k) + J_k \hat{d}_k\|$$

$$
\begin{aligned}
&= \|\lambda_k U_1(\Sigma_1^2+\lambda_k I)^{-1}U_1^T F(y_k) + \lambda_k U_2(\Sigma_2^2+\lambda_k I)^{-1}U_2^T F(y_k) + U_3 U_3^T F(y_k)\| \\
&\leq \lambda_k \|\Sigma_1^{-2}\| \|U_1 U_1^T F(y_k)\| + \|U_2 U_2^T F(y_k)\| + \|U_3 U_3^T F(y_k)\| \\
&\leq O\left(\|\bar{x}_k - x_k\|^{2+\nu}\right) + O\left(\|\bar{x}_k - x_k\|^{\nu+\gamma(1+\nu)}\right) \\
&\leq O\left(\|\bar{x}_k - x_k\|^{\nu+\gamma(1+\nu)}\right). \qquad (3.26)
\end{aligned}
$$

Hence, combining with (3.8), (3.25), (3.26), and Assumption 3.1, we know

$$
\begin{aligned}
c^{\frac{1}{\gamma}}\|\bar{x}_{k+1} - x_{k+1}\|^{\frac{1}{\gamma}} &\leq \|F(x_{k+1})\| = \|F(y_k + \alpha_k \hat{d}_k)\| \\
&\leq \|F(y_k) + \alpha_k J(y_k)\hat{d}_k\| + \frac{\kappa_{hj}}{1+\nu}\alpha_k^{1+\nu}\|\hat{d}_k\|^{1+\nu} \\
&\leq \|F(y_k) + \alpha_k J_k \hat{d}_k\| + \alpha_k\|(J(y_k) - J_k)\hat{d}_k\| + \frac{\kappa_{hj}}{1+\nu}\alpha_k^{1+\nu}\|\hat{d}_k\|^{1+\nu} \\
&\leq \|F(y_k) + \alpha_k J_k \hat{d}_k\| + k_{hj}\alpha_k\|\tilde{d}_k\|^{\nu}\|\hat{d}_k\| + \frac{\kappa_{hj}}{1+\nu}\alpha_k^{1+\nu}\|\hat{d}_k\|^{1+\nu} \\
&\leq O\left(\|\bar{x}_k - x_k\|^{\nu+\gamma(1+\nu)}\right) + O\left(\|\bar{x}_k - x_k\|^{\min\left\{1+2\nu, 3\nu+\gamma(1+\nu)-\frac{1}{\gamma}\right\}}\right) \\
&\quad + O\left(\|\bar{x}_k - x_k\|^{\min\left\{(1+\nu)^2, (1+\nu)(2\nu+\gamma(1+\nu)-\frac{1}{\gamma})\right\}}\right) \\
&\leq O\left(\|\bar{x}_k - x_k\|^{\xi}\right), \qquad (3.27)
\end{aligned}
$$

where $\xi = \min\left\{\nu+\gamma(1+\nu), 1+2\nu, 3\nu+\gamma(1+\nu)-\frac{1}{\gamma}, (1+\nu)^2, (1+\nu)(2\nu+\gamma(1+\nu)-\frac{1}{\gamma})\right\}$. Consider $\gamma \in (\frac{2}{3}, 1]$ and $\nu \in (\frac{1}{2}, 1]$, and we have

$$1 + 2\nu - (\nu + \gamma(1+\nu)) = (1-\gamma)(1+\nu) > 0,$$

and

$$(1+\nu)^2 - (1+2\nu) = \nu^2 > 0.$$

By $\nu > \frac{1}{2\gamma}$ and $\gamma \in (\frac{2}{3}, 1]$, we derive

$$3\nu + \gamma(1+\nu) - \frac{1}{\gamma} - (\nu + \gamma(1+\nu)) = 2\left(\nu - \frac{1}{2\gamma}\right) > 0,$$

and

$$(1+\nu)(2\nu+\gamma(1+\nu)-\frac{1}{\gamma}) - \left(3\nu+\gamma(1+\nu)-\frac{1}{\gamma}\right) = \nu\left(2\nu+\gamma+\nu\gamma-\frac{1}{\gamma}-1\right) > \nu\left(\gamma-\frac{1}{2}\right) > 0.$$

These mean that $\xi = \nu + \gamma(1+\nu)$ and $\{x_k\}$ converges to some solution of (1.1) with the rate of $\nu\gamma + \gamma^2(1+\nu)$.

Moreover, together with $\|\bar{x}_k - x_k\| \leq \|\bar{x}_{k+1} - x_k\| \leq \|\bar{x}_{k+1} - x_{k+1}\| + \|s_k\|$ and (3.27), we have

$$\|\bar{x}_k - x_k\| \leq 2\|s_k\|$$

for all sufficiently large k. It is clear from Lemma 3.1 that

$$\|s_{k+1}\| \leq O\left(\|s_k\|^{\nu\gamma+\gamma^2(1+\nu)}\right).$$

By the above explanation, along with the condition that $\nu > \max\left\{2\left(\frac{1}{\gamma}-1\right), \frac{1}{2\gamma}\right\}$, we can conclude that Theorem 3.1 is valid. The proof is complete. □

In addition, when the values of ν and γ are different, we have convergence rates as follows

$$\|s_{k+1}\| \leq \begin{cases} O\left(\|s_k\|^{\gamma+2\gamma^2}\right), & \text{if } \nu = 1; \\ O\left(\|s_k\|^{1+2\nu}\right), & \text{if } \gamma = 1; \\ O\left(\|s_k\|^{3}\right), & \text{if } \nu = 1 \text{ and } \gamma = 1. \end{cases}$$

4. Numerical experiments

This section shows the numerical results of the AATLM algorithm. All experiments were performed on a PC with an Intel i7-13700UH CPU with 32.00 GB RAM and MATLAB R2022a (64-bit).

We compare the AATLM algorithm with that of LM1 in [26], the MLM algorithm in [21], and the AMLM algorithm in [23]. Their parameters are chosen as follows:

LM1 : $q_0 = 10^{-4}, q_1 = 0.25, q_2 = 0.75, \mu_0 = 1, m_0 = 10^{-8}$;
MLM : $q_0 = 10^{-4}, q_1 = 0.25, q_2 = 0.75, \mu_0 = 1, m_0 = 10^{-8}, \delta = 1$;
AMLM : $q_0 = 10^{-4}, q_1 = 0.25, q_2 = 0.75, \mu_0 = 1, m_0 = 10^{-8}, \delta = 1, \hat{\alpha} = 4$;
AATLM : $q_0 = 10^{-4}, q_1 = 0.25, q_2 = 0.75, \theta = 0.6, \bar{\alpha}_0 = 1, \mu_0 = 1, m_0 = 10^{-8}, \tau = 0.1, a_1 = 4, a_2 = \frac{1}{4}$.

The termination condition of the algorithm is $\|J_k^T F_k\| \leq 10^{-6}$ or $k \geq 1000$. In the listed numerical results, "NF", "NJ", "NT = NF + NJ × n", "NK", and "Time" represent the numbers of functions, Jacobian evaluations, total evaluations, iterations, and CPU time, respectively. Examples 4.1 and 4.2 are two singular problems from [26]. These problems do not satisfy the local error bound condition, but satisfy the Hölderian local error bound condition. $J(x)$ of these problems are not Lipschitz continuous, but are Hölderian continuous.

Example 4.1. *[26] Consider the following Function 1:*

$$\begin{aligned} F_1(x) &= x_1 + 10x_2, \\ F_2(x) &= x_3 - x_4, \\ F_3(x) &= (x_2 - 2x_3)^{\frac{3}{2}}, \\ F_4(x) &= (x_1 - x_4)^{\frac{3}{2}}. \end{aligned}$$

The initial point is $x_0 = (3, 1, 0, 1)^T$, and the optimal solution is $x^* = (0, 0, 0, 0)^T$. The results are listed in Table 1.

Table 1. Numerical results of Example 4.1.

		LM1	MLM	AMLM	AATLM
(n, m)	x_0	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖
(4, 4)	$-10x_0$	11/11/55/10/0.00/7.99e−6	19/10/59/9/0.00/7.98e−6	17/9/53/8/0.00/4.62e−6	17/9/53/8/0.00/2.23e−6
	$-x_0$	10/10/50/9/0.00/4.14e−6	15/8/47/7/0.00/2.76e−6	13/7/41/6/0.00/8.90e−6	13/7/41/6/0.02/2.77e−6
	x_0	10/10/50/9/0.00/4.14e−6	15/8/47/7/0.00/2.76e−6	13/7/41/6/0.00/8.90e−6	13/7/41/6/0.00/2.77e−6
	$10x_0$	11/11/55/10/0.00/7.99e−6	19/10/59/9/0.00/3.15e−6	17/9/53/8/0.00/4.62e−6	17/9/53/8/0.00/1.63e−6
	$100x_0$	13/13/65/12/0.00/7.90e−6	23/12/71/11/0.00/6.77e−6	21/11/65/10/0.02/1.14e−5	17/9/53/8/0.00/1.59e−5

Example 4.2. *[26] Consider the following Function 2:*

$$\begin{aligned} F_1(x) &= x_1 + 10x_2, \\ F_2(x) &= x_3 - x_4, \\ F_3(x) &= (x_2 - 2x_3)^{\frac{4}{3}}, \\ F_4(x) &= (x_1 - x_4)^{\frac{4}{3}}. \end{aligned}$$

The initial point is $x_0 = (3, -1, 0, 1)^T$, and the optimal solution is $x^* = (0, 0, 0, 0)^T$. The results are listed in Table 2.

Table 2. Numerical results of Example 4.2.

		LM1	MLM	AMLM	AATLM
(n, m)	x_0	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖
(4, 4)	$-10x_0$	11/11/55/10/0.00/1.47e−6	19/10/59/9/0.00/7.46e−7	19/10/59/9/0.00/1.26e−6	15/8/47/7/0.00/1.75e−6
	$-x_0$	9/9/45/8/0.00/4.88e−6	15/8/47/7/0.00/7.98e−7	13/7/41/6/0.00/7.14e−6	13/7/41/6/0.00/2.44e−6
	x_0	9/9/45/8/0.00/4.88e−6	13/7/41/6/0.02/8.13e−6	13/7/41/6/0.02/2.30e−6	13/7/41/6/0.00/1.58e−6
	$10x_0$	11/11/55/10/0.00/1.47e−6	19/10/59/9/0.00/6.73e−7	17/9/53/8/0.00/1.27e−6	15/8/47/7/0.00/9.30e−7
	$100x_0$	12/12/60/11/0.00/2.67e−6	23/12/71/11/0.00/1.43e−6	21/11/65/10/0.00/2.31e−6	17/9/53/8/0.00/8.23e−7

Tables 1 and 2 show that the numbers of iterations and the function and Jacobian evaluations of the AATLM algorithm are less than that of the LM1, MLM, and AMLM algorithms. Due to the dimension of the problems being small, there is almost no difference in the CPU time.

Similar to [21, 23], we also consider the following singular problem [31]

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

where $F(x)$ is a nonsingular test function given by Moré, Garbow, and Hillstrom in [32], x^* is the root of $F(x)$, and $A \in \mathbb{R}^{n\times k}$ has full column rank with $1 \le k \le n$. There exists

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

with rank $n - k$. In this paper, we define

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

which means that the rank of $\hat{J}(x^*)$ is $n - 1$.

Example 4.3. *[32] Consider the extended Rosenbrock function*

$$\begin{aligned} F_{2i-1}(x) &= 10(x_{2i} - x_{i-1}^2), \\ F_{2i}(x) &= 1 - x_{2i-1}. \end{aligned}$$

The initial point is $x_0 = (-1.2, 1, -1.2, 1, ...)^T$, and the optimal solution is $x^* = (1, 1, ..., 1)^T$. The results are listed in Table 3.

Table 3. Numerical results of the extended Rosenbrock function.

		LM1	MLM	AMLM	AATLM
(n, m)	x_0	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖
(500, 500)	$-10x_0$	18/18/9018/17/1.56/2.50e−3	259/130/65259/129/18.61/2.23e−2	339/170/85339/169/53.36/2.73e−2	31/16/8031/15/4.38/1.59e−3
	$-x_0$	141/141/70641/140/16.06/3.56e−2	171/86/43171/85/7.77/2.02e−2	313/157/78813/156/27.23/2.84e−2	31/16/8031/15/4.67/1.38e−3
	x_0	55/55/27555/54/4.27/1.38e−2	223/112/56223/111/11.09/2.00e−2	337/169/84837/168/28.22/2.78e−2	101/51/25601/50/13.80/3.60e−1
	$10x_0$	21/21/10521/20/1.53/2.76e−3	187/94/47187/93/9.14/2.17e−2	339/170/85339/169/39.33/2.65e−2	31/16/8031/15/1.31/2.03e−3
	$100x_0$	24/24/12024/23/4.72/2.40e−3	63/32/16063/31/3.36/2.12e−3	379/190/95379/189/49.78/2.66e−2	35/18/9035/17/2.53/1.23e−3
(1000, 1000)	$-10x_0$	19/19/19019/18/11.58/1.89e−3	323/162/162323/161/224.11/3.48e−2	411/206/206411/205/245.09/4.56e−2	31/16/16031/15/28.98/2.37e−3
	$-x_0$	172/172/172172/171/88.95/5.66e−2	307/154/154307/153/152.39/3.46e−2	421/211/211421/210/333.70/4.51e−2	31/16/16031/15/36.92/2.10e−3
	x_0	62/62/62062/61/70.03/2.00e−2	321/161/161321/160/236.45/3.28e−2	447/224/224447/223/306.64/4.52e−2	181/91/91181/90/199.53/6.03e−2
	$10x_0$	22/22/22022/21/33.22/1.91e−3	305/153/153305/152/246.50/3.19e−2	455/228/228455/227/218.16/4.53e−2	31/16/16031/15/10.45/3.00e−3
	$100x_0$	25/25/25025/24/12.39/1.65e−3	61/31/31061/30/51.36/2.04e−3	55/28/28055/27/41.64/1.75e−3	35/18/18035/17/15.00/1.71e−3

Example 4.4. *[32] Consider the extended Powell singular function*

$$\begin{aligned} F_{4i-3}(x) &= x_{4i-3} + 10x_{4i-2}, \\ F_{4i-2}(x) &= 5^{1/2}(x_{4i-1} - x_{4i}), \\ F_{4i-1}(x) &= (x_{4i-2} - 2x_{4i-1})^2, \\ F_{4i}(x) &= 10^{1/2}(x_{4i-3} - x_{4i})^2. \end{aligned}$$

The initial point is $x_0 = (3, -1, 0, 1, ...)^T$, and the optimal solution is $x^* = (0, 0, ..., 0)^T$. The results are listed in Table 4.

Table 4. Numerical results of the extended Powell singular function.

		LM1	MLM	AMLM	AATLM
(n, m)	x_0	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖	NF/NJ/NT/NK/Time/‖F‖
(500, 500)	$-10x_0$	15/15/7515/14/0.95/5.29e−5	35/18/9035/17/1.09/1.77e−5	27/14/7027/13/0.11/1.51e−5	21/11/5521/10/0.14/4.31e−5
	$-x_0$	12/12/6012/11/0.25/3.80e−5	21/11/5521/10/0.48/5.28e−5	19/10/5019/9/0.09/3.35e−5	17/9/4517/8/0.11/2.45e−5
	x_0	12/12/6012/11/0.11/3.80e−5	21/11/5521/10/0.42/5.28e−5	19/10/5019/9/0.03/3.35e−5	17/9/4517/8/0.06/2.45e−5
	$10x_0$	15/15/7515/14/0.485.29e−5	35/18/9035/17/0.39/1.77e−5	27/14/7027/13/0.13/1.51e−5	21/11/5521/10/0.13/4.31e−5
	$100x_0$	19/19/9519/18/0.48/2.06e−5	29/15/7529/14/1.17/3.96e−5	45/23/11545/22/0.14/5.18e−5	27/14/7027/13/0.13/1.19e−5
(1000, 1000)	$-10x_0$	15/15/15015/14/0.59/2.37e−1	25/13/13025/12/0.89/4.42e−1	25/13/13025/12/1.53/1.94e−1	21/11/11021/10/0.81/2.52e−1
	$-x_0$	12/12/12012/11/0.58/3.80e−1	21/11/11021/10/0.45/3.60e−1	19/10/10019/9/1.19/3.80e−1	17/9/9017/8/0.75/3.73e−1
	x_0	12/12/12012/11/0.61/3.80e−1	21/11/11021/10/0.75/3.60e−1	19/10/10019/9/1.14/3.80e−1	17/9/9017/8/0.34/3.73e−1
	$10x_0$	15/15/15015/14/0.69/2.37e−1	25/13/13025/12/0.56/4.42e−1	25/13/13025/12/1.18/1.94e−1	21/11/11021/10/0.50/2.52e−1
	$100x_0$	19/19/19019/18/0.58/1.40e−1	31/16/16031/15/1.00/2.10e−1	31/16/16031/15/1.58/2.85e−1	27/14/14027/13/1.11/1.08e−1

Tables 3 and 4 show that the AATLM algorithm performs better than the LM1, MLM, and AMLM algorithms on the numbers of iterations and the function and Jacobian evaluations. For most problems, the AATLM algorithm has less CPU time than the other algorithms.

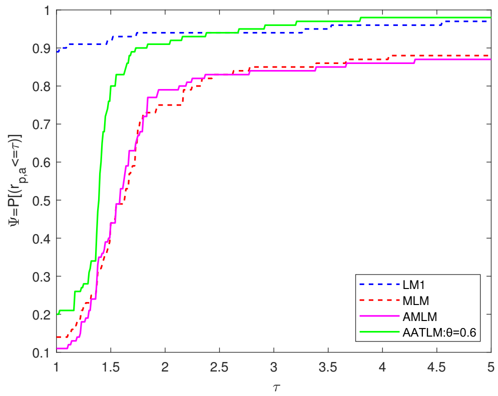
We tested 100 experiments and all functions are listed in Table 5. Problems 1 and 2 are Examples 4.1 and 4.2 from [26], Problems 3 and 4 are Examples 4.3 and 4.4, Problems 3–12 are from [32] and have the same form as (4.1), and Problems 13–16 are transformed from the CUTEr library in [33]. All of the test problems satisfy the assumptions required in this paper.

Table 5. Test functions.

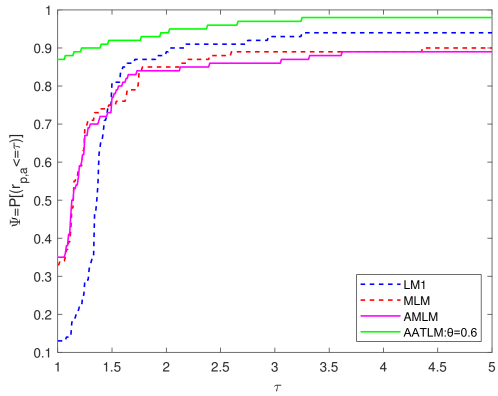
Prob.	Function	(n, m)	x_0	Prob.	Function	(n, m)	x_0
1	Function 1	(4, 4)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$	2	Function 2	(4, 4)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$
3	Extended Rosenbrock	(500, 500)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$	4	Extended Powell singular	(500, 500)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$
		(1000, 1000)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$			(1000, 1000)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$
5	Freudenstein and Roth	(2, 2)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$	6	Powell badly scaled	(2, 2)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$
7	Beale	(2, 3)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$	8	Helical valley	(3, 3)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$
9	Wood	(4, 6)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$	10	Extended Wood	(500, 750)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$
11	Trigonometric	(500, 500)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$	12	Brown almost-linear	(500, 500)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$
		(1000, 1000)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$			(1000, 1000)	$-10x_0, -x_0, x_0, 10x_0, 100x_0$
13	EG2	(500, 500)	$-x_0, x_0$	14	ARWHEAD	(500, 500)	$-10x_0, 10x_0$
		(1000, 1000)	$-x_0, x_0$			(1000, 1000)	$-10x_0, 10x_0$
15	LIARWHD	(500, 500)	$-10x_0, x_0, 10x_0$	16	TRIDIA	(500, 500)	$-10x_0, x_0, 10x_0$
		(1000, 1000)	$-10x_0, x_0, 10x_0$			(1000, 1000)	$-10x_0, x_0, 10x_0$

According to Dolan's [34] evaluation criteria, we show the performance profiles for the numbers of function evaluations, Jacobian evaluations, iterations, and CPU time of the algorithm in Figure 1. The parameter τ represents the performance ratio. When τ is close to 1 and Ψ remains constant, the numbers of Jacobian evaluations or iterations of the current algorithm are closer to the minimum value than the other algorithms. When τ is a constant and Ψ is close to 1, this means that the current algorithm can solve more problems.

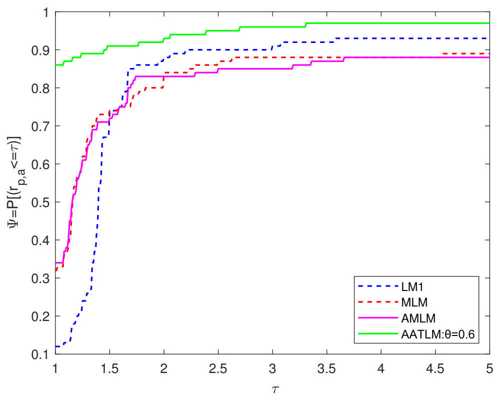
It can be seen from Figure 1 that the AATLM algorithm performs better than other algorithms in the numbers of Jacobian evaluations and iterations. From Figure 1(a), the AATLM algorithm performs better than the MLM and AMLM algorithms in the number of function evaluations. Since the LM1 algorithm calculates F_k only once in each iteration, the LM1 algorithm has a higher curve in Figure 1(a) when $\tau \in [1, 2.38]$. In Figure 1(b), the AATLM algorithm can solve more testing problems with less Jacobian evaluations. When $\tau \in (1.49, 5]$, the LM1 algorithm performs better than the MLM and AMLM algorithms. According to Figure 1(c), the AATLM can solve 86% of the problems with the least number of iterations, while the LM1, MLM, and AMLM can solve 12%, 32%, and 34% of the problems, respectively, which means that the AATLM algorithm could solve more problems with fewer iterations. In Figure 1(d), the LM1, MLM, AMLM, and AATLM algorithms can solve 60%, 54%, 42%, and 68% of the problems with the least CPU time, respectively. In summary, the results indicate that the AATLM algorithm is a promising method for solving nonlinear equations.



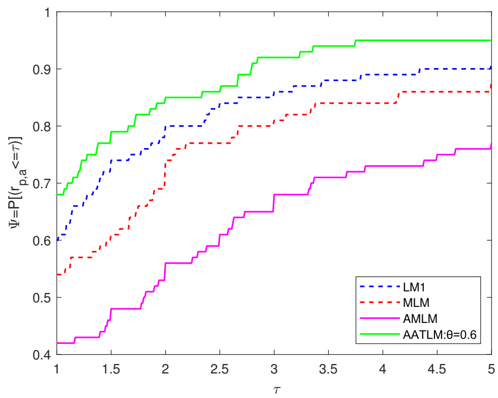
(a) Performance profiles for the number of function evaluations.



(b) Performance profiles for the number of Jacobian evaluations.



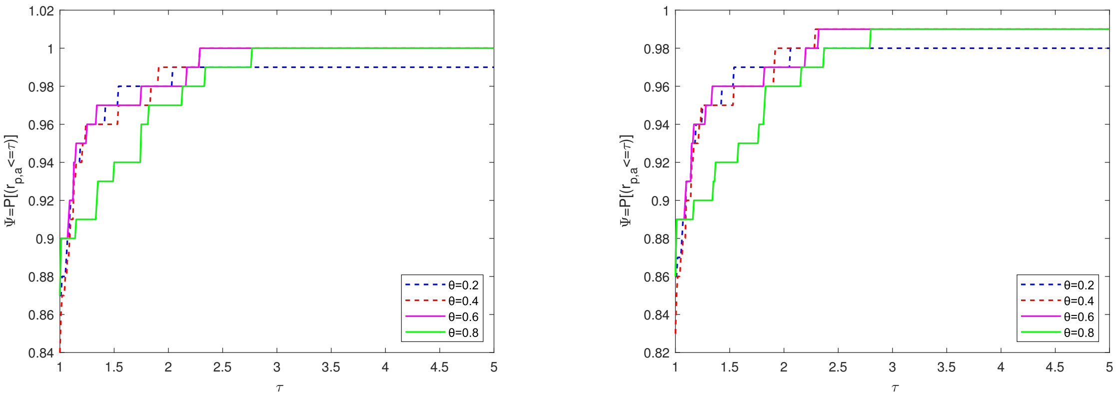
(c) Performance profiles for the number of iterations.



(d) Performance profiles for the CPU time.

Figure 1. Performance profiles of the numerical results.

In addition, we also consider the influence of different θ on the AATLM algorithm. We show the performance profiles for the numbers of Jacobian evaluations and the iterations of the AATLM algorithm in Figure 2, where θ is chosen from the set $\{0.2, 0.4, 0.6, 0.8\}$. We find that when θ in the AATLM algorithm is 0.6, the curve is higher than the others. This means that the new algorithm with $\theta = 0.6$ can solve more problems with fewer Jacobian evaluations and iterations.



(a) Performance profiles for the number of Jacobian evaluations.

(b) Performance profiles for the number of iterations.

Figure 2. Performance profiles of the numerical results with different θ.

5. Conclusions

In this paper, we constructed a new LM parameter in the form of a convex combination to obtain the LM step and the approximate step. A new modified Metropolis criterion was introduced to update the upper bound of the approximate step size, so as to obtain an adaptive acceleration two-step LM algorithm. The global and local convergence of the new algorithm were studied under the Hölderian local error bound condition and the Hölderian continuity of the Jacobian, which are more general than the local error bound condition and the Lipschitz continuity of the Jacobian. The numerical results showed the efficiency of the AATLM algorithm. In the course of research, we noticed that different LM parameters could be considered at different stages of the algorithm. In future work, we will explore a new LM parameter and introduce a nonmonotone technique into the two-step LM algorithm to solve nonlinear equations.

Author contributions

Dingyu Zhu: conceptualization, writing–original draft, software, methodology, writing–review and editing; Yueting Yang: writing–original draft, supervision, funding acquisition, methodology, project administration, writing–review and editing; Mingyuan Cao: writing–original draft, supervision, funding acquisition, methodology, project administration, writing–review and editing. All authors have read and approved the final version of the manuscript for publication.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors would like to thank the Science and Technology Development Planning Grant Program of Jilin Province (YDZJ202101ZYTS167, 20200403193SF) and the Graduate Innovation Project of Beihua University (2023005, 2023035).

Conflict of interest

The authors declare no conflicts of interest.

References

1. G. D. A. Moura, S. D. T. M. Bezerra, H. P. Gomes, S. A. D. Silva, Neural network using the Levenberg–Marquardt algorithm for optimal real-time operation of water distribution systems, *Urban Water J.*, **15** (2018), 692–699. https://doi.org/10.1080/1573062X.2018.1539503
2. Y. J. Sun, P. P. Wang, T. T. Zhang, K. Li, F. Peng, C. G. Zhu, Principle and performance analysis of the Levenberg–Marquardt algorithm in WMS spectral line fitting, *Photonics*, **9** (2022), 999. https://doi.org/10.3390/photonics9120999
3. A. Alloqmani, O. Alsaedi, N. Bahatheg, R. Alnanih, L. Elrefaei, Design principles-based interactive learning tool for solving nonlinear equations, *Comput. Syst. Sci. Eng.*, **40** (2022), 1023–1042. https://doi.org/10.32604/csse.2022.019704
4. Z. W. Liao, F. Y. Zhu, W. Y. Gong, S. J. Li, X. Y. Mi, AGSDE: Archive guided speciation-based differential evolution for nonlinear equations, *Appl. Soft Comput.*, **122** (2022), 108818. https://doi.org/10.1016/j.asoc.2022.108818
5. Z. Seifi, A. Ghorbani, A. Abdipour, Time-domain analysis and experimental investigation of electromagnetic wave coupling to RF/microwave nonlinear circuits, *J. Electromagnet Wave.*, **35** (2021), 51–70. https://doi.org/10.1080/09205071.2020.1825994
6. A. Rothwell, Numerical methods for unconstrained optimization, In: *Optimization methods in structural design*, Cham: Springer, 2017, 83–106. https://doi.org/10.1007/978-3-319-55197-5
7. G. L. Yuan, M. J. Zhang, A three-terms Polak-Ribière-Polyak conjugate gradient algorithm for large-scale nonlinear equations, *J. Comput. Appl. Math.*, **286** (2015), 186–195. https://doi.org/10.1016/j.cam.2015.03.014
8. G. L. Yuan, Z. X. Wei, X. W. Lu, A BFGS trust-region method for nonlinear equations, *Computing*, **92** (2011), 317–333. https://doi.org/10.1007/s00607-011-0146-z
9. J. H. Zhang, Y. Q. Wang, J. Zhao, On maximum residual nonlinear Kaczmarz-type algorithms for large nonlinear systems of equations, *J. Comput. Appl. Math.*, **425** (2023), 115065. https://doi.org/10.1016/j.cam.2023.115065
10. J. N. Wang, X. Wang, L. W. Zhang, Stochastic regularized Newton methods for nonlinear equations, *J. Sci. Comput.*, **94** (2023), 51. https://doi.org/10.1007/s10915-023-02099-4
11. R. Behling, D. S. Gonçalves, S. A. Santos, Local convergence analysis of the Levenberg–Marquardt framework for Nonzero–Residue nonlinear least-squares problems under an error bound condition, *J. Optim. Theory Appl.*, **183** (2019), 1099–1122. https://doi.org/10.1007/s10957-019-01586-9

12. E. H. Bergou, Y. Diouane, V. Kungurtsev, Convergence and complexity analysis of a Levenberg–Marquardt algorithm for inverse problems, *J. Optim. Theory Appl.*, **185** (2020), 927–944. https://doi.org/10.1007/s10957-020-01666-1

13. K. Levenberg, A method for the solution of certain non-linear problems in least squares, *Quart. Appl. Math.*, **2** (1944), 164–168. https://doi.org/10.1090/qam/10666

14. D. W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.*, **11** (1963), 431–441. https://doi.org/10.1137/0111030

15. N. Yamashita, M. Fukushima, On the rate of convergence of the Levenberg–Marquardt method, In: *Topics in numerical analysis, computing supplementa*, Vienna: Springer, 2001, 239–249. https://doi.org/10.1007/978-3-7091-6217-0_18

16. J. Y. Fan, Y. X. Yuan, On the convergence of a new Levenberg–Marquardt method, Report, Institute of Computational Mathematics and Scientific/Engineering Computing, Beijing: Chinese Academy of Science, 2001.

17. J. Y. Fan, A Modified Levenberg–Marquardt algorithm for singular system of nonlinear equations, *J. Comput. Math.*, **21** (2003), 625–636.

18. K. Amini, F. Rostami, G. Caristi, An efficient Levenberg–Marquardt method with a new LM parameter for systems of nonlinear equations, *Optimization*, **67** (2018), 637–650. https://doi.org/10.1080/02331934.2018.1435655

19. C. F. Ma, L. H. Jiang, Some research on Levenberg–Marquardt method for the nonlinear equations, *Appl. Math. Comput.*, **184** (2007), 1032–1040. https://doi.org/10.1016/j.amc.2006.07.004

20. J. Y. Fan, J. Y. Pan, A note on the Levenberg–Marquardt parameter, *Appl. Math. Comput.*, **207** (2009), 351–359. https://doi.org/10.1016/j.amc.2008.10.056

21. J. Y. Fan, The modified Levenberg–Marquardt method for nonlinear equations with cubic convergence, *Math. Comp.*, **81** (2012), 447–466. https://doi.org/10.1090/S0025-5718-2011-02496-8

22. J. Y. Fan, J. L. Zeng, A Levenberg–Marquardt algorithm with correction for singular system of nonlinear equations, *Appl. Math. Comput.*, **219** (2013), 9438–9446. https://doi.org/10.1016/j.amc.2013.03.026

23. J. Y. Fan, Accelerating the modified Levenberg–Marquardt method for nonlinear equations, *Math. Comp.*, **83** (2014), 1173–1187. https://doi.org/10.1090/S0025-5718-2013-02752-4

24. X. D. Zhu, G. H. Lin, Improved convergence results for a modified Levenberg–Marquardt method for nonlinear equations and applications in MPCC, *Optim. Method. Softw.*, **31** (2016), 791–804. https://doi.org/10.1080/10556788.2016.1171863

25. H. Y. Wang, J. Y. Fan, Convergence rate of the Levenberg–Marquardt method under Hölderian local error bound, *Optim. Method. Softw.*, **35** (2020), 767–786. https://doi.org/10.1080/10556788.2019.1694927

26. M. L. Zeng, G. H. Zhou, Improved convergence results of an efficient Levenberg–Marquardt method for nonlinear equations, *J. Appl. Math. Comput.*, **68** (2022), 3655–3671. https://doi.org/10.1007/s12190-021-01599-6

27. L. Chen, Y. F. Ma, A modified Levenberg–Marquardt method for solving system of nonlinear equations, *J. Appl. Math. Comput.*, **69** (2023), 2019–2040. https://doi.org/10.1007/s12190-022-01823-x

28. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.*, **21** (1953), 1087–1092. https://doi.org/10.1063/1.1699114

29. R. Behling, A. Iusem, The effect of calmness on the solution set of systems of nonlinear equations, *Math. Program.*, **137** (2013), 155–165. https://doi.org/10.1007/s10107-011-0486-7

30. G. W. Stewart, J. G. Sun, *Matrix perturbation theory*, New York: Academic Press, 1990.

31. R. B. Schnabel, P. D. Frank, Tensor methods for nonlinear equations, *SIAM J. Numer. Anal.*, **21** (1984), 815–843. https://doi.org/10.1137/0721054

32. J. J. Moré, B. S. Garbow, K. E. Hillstrom, Testing unconstrained optimization software, *ACM T. Math. Software*, **7** (1981), 17–41. https://doi.org/10.1145/355934.355936

33. N. I. M. Gould, D. Orban, P. L. Toint. CUTEr and SifDec: A constrained and unconstrained testing environment, revisited, *ACM T. Math. Software*, **29** (2003), 373–394. https://doi.org/10.1145/962437.962439

34. E. D. Dolan, J. J. Moré, Benchmarking optimization software with performance profiles, *Math. Program.*, **91** (2002), 201–213. https://doi.org/10.1007/s101070100263