

一种改进的非单调信赖域两步 Levenberg-Marquardt方法

何素丹

长沙理工大学数学与统计学院, 湖南 长沙

收稿日期: 2025年5月25日; 录用日期: 2025年6月17日; 发布日期: 2025年6月27日

摘要

本文提出了一种求解非线性方程组的改进型非单调信赖域两步Levenberg-Marquardt方法, 证明了该方法在适当条件下具有全局收敛性。数值试验表明该方法比单步Levenberg-Marquardt方法具有更好的数值效果。

关键词

非线性方程组, Levenberg-Marquardt方法, 非单调信赖域, 全局收敛

An Improved Non-Monotonic Trust Region Two-Step Levenberg-Marquardt Method

Sudan He

School of Mathematics and Statistics, Changsha University of Science & Technology, Changsha Hunan

Received: May 25th, 2025; accepted: Jun. 17th, 2025; published: Jun. 27th, 2025

Abstract

In this paper, an improved nonmonotonic trust region two-step Levenberg-Marquardt method for solving nonlinear equations is proposed, and it is proved that the method has global convergence under appropriate conditions. Numerical experiments show that the method has better numerical results than the one-step Levenberg-Marquardt method.

Keywords

Nonlinear Equations, Levenberg-Marquardt Method, Non-Monotonic Trust Region, Global

Convergence

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

本文考虑求解如下非线性方程组的数值方法:

$$F(x) = 0, \quad (1.1)$$

其中 F 是 R^n 到 R^n 的连续可微函数, 其 Jacobian 阵记为 $J(x) \triangleq F'(x)$ 。非线性方程组在物理学、经济学和产品管理等许多领域都具有广泛的应用, 例如非线性优化问题的 KKT 系统、非线性微分方程的离散形式都是非线性方程组[1] [2]。因此设计高效求解非线性方程组的算法是计算数学领域的一个重要研究课题。

有许多实用的算法可以求解非线性方程组, 其中 Levenberg-Marquardt (LM)方法是一种经典方法。在每次迭代中, 它都会计算 LM 方向

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

其中 $F_k = F(x_k)$, $J_k = F'(x_k)$, 以及 LM 参数 $\lambda_k > 0$ 在每次迭代中更新。

LM 参数 λ_k 的选择对 LM 方法的计算效率至关重要。为了提高计算效率, 许多学者对 LM 参数进行研究。Yamashita 和 Fukushima [3]证明, 当 $\lambda_k = \|F_k\|^2$ 时, LM 方法在局部误差界这一较弱的条件下具有二次收敛速度。然而, 当序列 $\{x_k\}$ 远离解集 X 时, $\|F_k\|$ 可能非常大, 会导致 λ_k 也很大, 这将使得 d_k 变得很小, 从而降低算法的效率。于是 Fan 和 Yuan [4]使用了 $\lambda_k = \mu_k \|F_k\|^\delta$, 其中 $\delta \in (1, 2]$, μ_k 在每次迭代中都使用信赖域方法更新, 此时 LM 方法在一些合适的条件下也具有二次收敛速度。进一步, Ma 和 Jiang [5]将 $\|J_k^T F_k\|$ 引入 LM 方法用于求解奇异非线性方程组, 使用了 $\|F_k\|$ 和 $\|J_k^T F_k\|$ 的凸组合作为新的 LM 参数 $\lambda_k = \theta \|F_k\| + (1-\theta) \|J_k^T F_k\|$, 其中 $\theta \in [0, 1]$ 。随后, Fan 和 Pan [6]提出了 $\lambda_k = \mu_k [\theta \|F_k\| + (1-\theta) \|J_k^T F_k\|]$, 并证明了在弱于非奇异性的局部误差条件下仍然保持二次收敛。

为了提高算法的收敛速度和效率, Fan [7]提出了一种两步 LM 方法。在每次迭代中, 首先通过求解

$$\left(J_k^T J_k + \lambda_k I\right) d = -J_k^T F_k, \quad \text{其中 } \lambda_k = \mu_k \|F_k\|^\delta, \quad \delta \in [1, 2], \quad (1.2)$$

来获得 LM 方向 $d_k^{LM} = -\left(J_k^T J_k + \lambda_k I\right)^{-1} J_k^T F_k$, 其中 μ_k 由信赖域方法更新。再通过求解

$$\left(J_k^T J_k + \lambda_k I\right) d = -J_k^T F(y_k), \quad \text{其中 } y_k = x_k + d_k^{LM}, \quad (1.3)$$

来获得近似的 LM 方向 d_k^{MLM} , 并设置迭代方向为 $d_k = d_k^{LM} + d_k^{MLM}$ 。Fan 证明了该两步 LM 方法具有全局收敛性和在局部误差界下具有三次收敛速度。

近年来, 非单调方法被应用于多种优化算法中, 显著提升了算法的收敛性能。1986 年, Grippo 等[8]提出了非单调 Armijo 线搜索, 该方法是非线性优化的非单调算法的新突破。随后, Deng 等[9]首次将非单调思想与信赖域方法相结合, 提出了一种具有强收敛性的非单调信赖域方法。研究表明, 使用非单调技术可以提高找到全局最优值的可能性和算法的收敛率[8] [10]。由于非单调信赖域方法的良好数值效果, 越来越多的学者致力于对非单调方法的研究。最近, Zhao 等[11]在 LM 方法中采用非单调信赖域方法来确保全局收敛。传统的信赖域方法要求目标函数值 $\|F_k\|^2$ 严格单调下降, 这有可能会陷入局部最优, 而文

献[11]通过引入 W_{k+1} :

$$W_{k+1} = (1-\tau)W_k + \tau\|F_{k+1}\|^2, \text{ 其中 } \tau \in (0,1],$$

来允许 $\|F_{k+1}\|^2$ 偶尔上升, 只要 W_{k+1} 整体下降即可, 从而提升算法的全局收敛性。并通过与 Fan [12]的单调信赖域 LM 方法进行数值实验比较, 表明引入的非单调信赖域方法提高了 LM 方法的数值性能。

Zhao 在文献[11]中还利用 $\|F_k\|$ 和 $\|J_k^T F_k\|$ 的凸组合, 提出了一种新的 LM 参数 $\lambda_k = \mu_k \left[\theta \|F_k\|^\delta + (1-\theta) \|J_k^T F_k\|^\delta \right]$, 其中 $\theta \in [0,1]$ 和 $\delta \in (0,3)$ 。这是 LM 参数的更通用选择, 它不仅包括上述所提到的所有 LM 选择, 而且在局部误差界下, 证明了新的 LM 方法至少具有阶数为 $\min\{1+\delta, 4-\delta, 2\}$ 的超线性收敛率。

受上述文献的启发, 在本文中, 我们旨在提出一种新的 LM 方法, 该方法采用文献[11]中的 LM 参数以及采用非单调信赖域, 将其改进为两步 LM 方法。在文献[7]中两步 LM 方法已证明具有三次收敛速度, 通过与[7]同样的分析, 可以证明本文改进的两步 LM 方法在 $\delta \in [1,2]$ 时也具有三次收敛速度。

本文的组织结构如下。在第 2 节中, 我们详细介绍改进型非单调信赖域的两步 LM 方法。在第 3 节中, 建立了所提方法的全局收敛性。在第 4 节中, 我们应用新的两步 LM 方法求解一些非线性方程并报告数值结果。最后, 我们在第 5 节进行总结。在整篇论文中, 假设 X 表示(1.1)的解集是非空的, 且在所有情况下 $\|\cdot\|$ 是指 2-范数。

2. 算法的提出

我们取

$$\Phi(x) = \|F(x)\|^2$$

作为(1.1)的评价函数。两步 LM 方法在每一次迭代中不仅要通过(1.2)计算 LM 方向 d_k , 还要通过(1.3)计算近似 LM 方向 \hat{d}_k 。

在标准的信赖域方法中, 通常我们会将 $\Phi(x)$ 在第 k 次迭代时的实际下降量 $Ared_k$ 、预测下降量为 $Pred_k$ 和下降比 r_k 分别定义为

$$Ared_k = \|F_k\|^2 - \|F(x_k + d_k + \hat{d}_k)\|^2, \quad (2.1)$$

$$Pred_k = \|F_k\|^2 - \|F_k + J_k(d_k + \hat{d}_k)\|^2, \quad (2.2)$$

$$r_k = \frac{Ared_k}{Pred_k}. \quad (2.3)$$

但是在两步 LM 方法中, 因为不能证明(2.2)是非负的, 所以我们不能像往常一样将预测下降量定义为(2.2), 因此需要开发一种新的预测下降量。

注意到 LM 方向 d_k 是下列最小化关于 d 的优化问题的解:

$$\min_{d \in \mathbb{R}^n} \varphi_{k,1}(d) \triangleq \|F_k + J_k d\|^2 + \lambda_k \|d\|^2. \quad (2.4)$$

如果我们令

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

那么可以验证 d_k 也是下列信赖域问题的解:

$$\min_{d \in R^n} \|F_k + J_k d\|^2 \quad \text{s.t. } \|d\| \leq \Delta_{k,1}.$$

根据 Powell [13]给出的著名结果, 我们知道

$$\|F_k\|^2 - \|F_k + J_k d_k\|^2 \geq \|J_k^T F_k\| \min \left\{ \|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|} \right\}. \quad (2.5)$$

同理, 近似 LM 方向 \hat{d}_k 不仅是下列最小化关于 d 的优化问题的解:

$$\min_{d \in R^n} \varphi_{k,2}(d) \triangleq \|F(y_k) + J_k d\|^2 + \lambda_k \|d\|^2. \quad (2.6)$$

也是下列信赖域问题的解:

$$\min_{d \in R^n} \|F(y_k) + J_k d\|^2 \quad \text{s.t. } \|d\| \leq \Delta_{k,2}.$$

其中 $\Delta_{k,2} = \|\hat{d}_k\| = \left\| -(J_k^T J_k + \lambda_k I)^{-1} J_k^T F(y_k) \right\|$ 。同样可以得到

$$\|F(y_k)\|^2 - \|F(y_k) + J_k \hat{d}_k\|^2 \geq \|J_k^T F(y_k)\| \min \left\{ \|\hat{d}_k\|, \frac{\|J_k^T F(y_k)\|}{\|J_k^T J_k\|} \right\}. \quad (2.7)$$

现在, 根据不等式(2.5)和(2.7), 我们可以合理地将新的预测下降量定义为

$$\widehat{Pred}_k = \|F_k\|^2 - \|F_k + J_k d_k\|^2 + \|F(y_k)\|^2 - \|F(y_k) + J_k \hat{d}_k\|^2. \quad (2.8)$$

并且它总是非负的, 满足以下不等式

$$\widehat{Pred}_k \geq \|J_k^T F_k\| \min \left\{ \|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|} \right\} + \|J_k^T F(y_k)\| \min \left\{ \|\hat{d}_k\|, \frac{\|J_k^T F(y_k)\|}{\|J_k^T J_k\|} \right\}. \quad (2.9)$$

受文献[11]的非单调信赖域方法的启发, 我们定义新的实际下降量为

$$\widehat{Ared}_k = W_k - \|F(x_k + d_k + \hat{d}_k)\|^2 \quad (2.10)$$

其中 $W_{k+1} = (1-\tau)W_k + \tau\|F_{k+1}\|^2$, $\tau \in (0,1]$ 。值得注意的是, W_k 是 W_{k-1} 和 $\|F_k\|^2$ 的凸组合。由于我们设定 $W_0 = \|F_0\|^2$, 因此 W_k 是 $\|F_0\|^2, \|F_1\|^2, \dots, \|F_k\|^2$ 的凸组合。从而可以定义新的下降比为 $\hat{r}_k = \frac{\widehat{Ared}_k}{\widehat{Pred}_k}$, 用于决定是

否接受试探步以及如何调整 LM 参数。

接下来, 我们将详细地描述新的两步 LM 方法。

算法 1

步 1. 选择一个初始点 $x_0 \in R^n$, 令 $W_0 = \|F_0\|^2$, 给定参数 $\mu_0 \geq m_0 > 0$, $0 < p_0 \leq p_1 \leq p_2 < 1$, $\theta \in [0,1]$, $\tau \in (0,1]$ 和 $\delta \in (0,3)$ 。置 $k := 0$ 。

步 2. 如果 $\|J_k^T F_k\| = 0$, 则停止计算。否则, 令

$$\lambda_k = \mu_k \left[(1-\theta)\|F_k\|^\delta + \theta\|J_k^T F_k\|^\delta \right], \quad (2.11)$$

通过求解以下线性方程组, 得到 d_k :

$$(J_k^T J_k + \lambda_k I) d = -J_k^T F_k, \quad (2.12)$$

令 $y_k = x_k + d_k$; 再求解以下方程组, 得到 \hat{d}_k :

$$(J_k^T J_k + \lambda_k I)d = -J_k^T F(y_k), \quad (2.13)$$

再令 $s_k = d_k + \hat{d}_k$ 。

步 3. 通过(2.8)和(2.10)计算下降比

$$\hat{r}_k = \frac{\widehat{Ared}_k}{\widehat{Pred}_k} \quad (2.14)$$

步 4. 令

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{若 } \hat{r}_k \geq p_0, \\ x_k, & \text{若 } \hat{r}_k < p_0. \end{cases} \quad (2.15)$$

令

$$W_{k+1} = (1-\tau)W_k + \tau \|F_{k+1}\|^2. \quad (2.16)$$

步 5. 选择 μ_{k+1} 为

$$\mu_{k+1} = \begin{cases} 4\mu_k, & \text{若 } \hat{r} < p_1, \\ \mu_k, & \text{若 } \hat{r} \in [p_1, p_2], \\ \max\left\{\frac{\mu_k}{4}, m_0\right\}, & \text{若 } \hat{r} > p_2. \end{cases} \quad (2.17)$$

置 $k := k+1$, 转步 2。

注:

两步 LM 方法的计算成本与单步 LM 方法的计算成本几乎相同, 因为(2.13)只涉及 $F(y_k)$, 并且在求解(2.12)之后, 可以利用 $J_k^T J_k + \lambda_k I$ 的可用分解。

3. 全局收敛性

为了证明算法 1 的全局收敛性, 我们做出如下假设。

假设 A

- (1) $F(x)$ 在 R^n 上是连续可微的。
- (2) $F(x)$ 和 $J(x)$ 在 R^n 上是 Lipschitz 连续的, 即对 $\forall x, y \in R^n$, 存在正的常数 L , 使得

$$\|J(x) - J(y)\| \leq L\|x - y\|, \quad (3.1)$$

$$\|F(x) - F(y)\| \leq L\|x - y\|. \quad (3.2)$$

假设 A 也意味着, 对 $\forall x, y \in R^n$,

$$\|F(y) - F(x) - J(x)(y-x)\| \leq L\|y-x\|^2. \quad (3.3)$$

$$\|J(x)\| \leq L. \quad (3.4)$$

引理 3.1 设 $\{x_k\}$ 是由算法 1 产生的序列, 则对所有的 $k \geq 0$, 都有

$$\|F_k\|^2 \leq W_k, \quad W_{k+1} \leq W_k \quad \text{和} \quad \|F_k\| \leq \|F_0\|.$$

证明. 首先, 我们假设对某些 k , 有 $\|F_k\|^2 \leq W_k$ 。如果 $\hat{r}_k < p_0$, 那么由(2.15)可得 $x_{k+1} = x_k$, 则

$$W_k \geq \|F_k\|^2 = \|F_{k+1}\|^2. \quad (3.5)$$

如果 $\widehat{r}_k \geq p_0$, 那么由(2.15)可得 $x_{k+1} = x_k + s_k$, 再由(2.14)得

$$\widehat{r}_k = \frac{W_k - \|F(x_k + s_k)\|^2}{\widehat{Pred}_k} = \frac{W_k - \|F_{k+1}\|^2}{\widehat{Pred}_k} \geq p_0,$$

上式结合(2.9), 意味着

$$W_k \geq \|F_{k+1}\|^2 + p_0 \widehat{Pred}_k \geq \|F_{k+1}\|^2. \quad (3.6)$$

于是由(2.16)可知

$$W_{k+1} = (1-\tau)W_k + \tau \|F_{k+1}\|^2 \geq \|F_{k+1}\|^2.$$

因为 $W_0 = \|F_0\|^2$, 所以通过对 k 进行归纳, 我们可以得到对所有的 $k \geq 0$, 都有

$$\|F_k\|^2 \leq W_k,$$

这就证得了引理的第一个结果。

然后, 结合(3.5)和(3.6)可知对所有的 $k \geq 0$, $\|F_{k+1}\|^2 \leq W_k$, 再由(2.16)可得

$$W_{k+1} = (1-\tau)W_k + \tau \|F_{k+1}\|^2 \leq (1-\tau)W_k + \tau W_k = W_k, \text{ 即对所有的 } k \geq 0, \text{ 都有}$$

$$W_{k+1} \leq W_k,$$

这就证得了引理的第二个结果。

此外, 对所有的 $k \geq 0$, 我们有

$$\|F_k\|^2 \leq W_k \leq W_0 = \|F_0\|^2.$$

这就证得了引理的第三个结果。 □

定理 3.1 设 $\{x_k\}$ 是由算法 1 产生的序列, 若假设 A 成立, 则有

$$\liminf_{k \rightarrow \infty} \|J_k^T F_k(x_k)\| = 0. \quad (3.7)$$

证明. 用反证法。假设 $\liminf_{k \rightarrow \infty} \|J_k^T F_k\| \neq 0$, 则存在常数 $\bar{k} > 0, \eta > 0$, 使得对 $\forall k > \bar{k}$, 有

$$\|J_k^T F_k\| \geq \eta, \quad (3.8)$$

定义包含所有成功迭代点的指标集为

$$S = \{k \mid \widehat{r}_k \geq p_0\}$$

我们考虑以下两种情况, 并得出矛盾。

情形 1. S 是无限的。在这种情形下, 由(2.9), (2.15), (2.16), (3.4)和(3.8)可得, 对所有的 $k \in S$ 以及 $k > \bar{k}$, 有

$$\begin{aligned} W_k - \|F_{k+1}\|^2 &= W_k - \|F(x_k + s_k)\|^2 = \widehat{Pred}_k \\ &\geq p_0 \widehat{Pred}_k \\ &\geq p_0 \|J_k^T F_k\| \min \left\{ \|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|} \right\} + p_0 \|J_k^T F(y_k)\| \min \left\{ \|\hat{d}_k\|, \frac{\|J_k^T F(y_k)\|}{\|J_k^T J_k\|} \right\}, \\ &\geq p_0 \eta \min \left\{ \|d_k\|, \frac{\eta}{L} \right\}. \end{aligned}$$

根据引理 3.1 的第二个结果, 序列 $\{W_k\}$ 单调递减且有下界, 即存在一个常数 $W^* \geq 0$, 使得 $\lim_{k \rightarrow \infty} W_k = W^*$ 。再由(2.16), 我们有

$$\lim_{k \rightarrow \infty} \|F_k\|^2 = \lim_{k \rightarrow \infty} \frac{W_k - (1-\tau)W_{k-1}}{\tau} = W^*.$$

于是根据 $\lim_{k \rightarrow \infty} \|F_k\| = \lim_{k \rightarrow \infty} W_k = W^*$, 可知 $\lim_{k \in S, k \rightarrow \infty} d_k = 0$ 。

然而当 $k \notin S$ 时, 有 $x_{k+1} - x_k = 0$ 。因此可以得到

$$\lim_{k \rightarrow \infty} d_k = 0.$$

上式结合(2.12)和(3.8)可得

$$\lim_{k \rightarrow \infty} \lambda_k = +\infty. \quad (3.9)$$

根据引理 3.1 的第三个结果以及(3.4), 有

$$(1-\theta)\|F_k\|^\delta + \theta\|J_k^T F_k\|^\delta \leq (1-\theta)\|F_0\|^\delta + \theta L^\delta \|F_0\|^\delta.$$

上式结合(2.11)和(3.9)可得

$$\lim_{k \rightarrow \infty} \mu_k = +\infty. \quad (3.10)$$

此外, 由 \hat{d} 的定义和(3.3), 对所有充分大的 k 有

$$\begin{aligned} \|\hat{d}_k\| &= \left\| -\left(J_k^T J_k + \lambda_k I\right)^{-1} J_k^T F(y_k) \right\| \\ &\leq \left\| \left(J_k^T J_k + \lambda_k I\right)^{-1} J_k^T \left(F(y_k) - F_k - J_k d_k\right) \right\| \\ &\quad + \left\| \left(J_k^T J_k + \lambda_k I\right)^{-1} J_k^T F_k \right\| + \left\| \left(J_k^T J_k + \lambda_k I\right)^{-1} J_k^T J_k d_k \right\| \\ &\leq \left\| \left(J_k^T J_k + \lambda_k I\right)^{-1} \right\| \left\| J_k \right\| \left\| F(y_k) - F_k - J_k d_k \right\| + \|d_k\| + \|d_k\| \\ &\leq \lambda_k^{-1} L^2 \|d_k\|^2 + 2\|d_k\| \\ &\leq \bar{c} \|d_k\|. \end{aligned} \quad (3.11)$$

其中 \bar{c} 是某个正的常数。于是

$$\|s_k\| = \|d_k + \hat{d}_k\| \leq (1 + \bar{c}) \|d_k\|.$$

故由(2.1), (2.8), (2.9), (3.4), (3.8)和上式, 对 $\forall k > \bar{k}$,

$$\begin{aligned} \left| \frac{\widehat{Ared}_k - \widehat{Pred}_k}{\widehat{Pred}_k} \right| &\leq \frac{\left| \|F(x_k + s_k)\|^2 - \|F_k + J_k d_k\|^2 + \|F(y_k)\|^2 - \|F(y_k) + J_k \hat{d}_k\|^2 \right|}{\left\| J_k^T F_k \right\| \min \left\{ \|d_k\|, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|} \right\} + \|J_k^T F(y_k)\| \min \left\{ \|\hat{d}_k\|, \frac{\|J_k^T F(y_k)\|}{\|J_k^T J_k\|} \right\}} \\ &\leq \frac{\|F_k + J_k s_k\| \mathcal{O}(\|s_k\|^2 + \|d_k\|^2) + \mathcal{O}(\|s_k\|^4 + \|d_k\|^4) + \|F_k + J_k d_k\| \mathcal{O}(\|d_k\|^2)}{\eta \min \left\{ \|d_k\|, \frac{\eta}{L^2} \right\}} \\ &\leq \frac{\|F_k\| \mathcal{O}(\|d_k\|^2) + \|J_k \hat{d}_k\| \mathcal{O}(\|d_k\|^2)}{\|d_k\|} \\ &\rightarrow 0. \end{aligned}$$

即得 $k \rightarrow \infty$, 有 $\frac{Ared_k}{Pred_k} \rightarrow 1$ 。根据引理 3.1 的第一个结果, 对 $\forall k \geq 0$, 都有 $W_k \geq \|F_k\|^2$, 于是

$\widehat{Ared}_k \geq Ared_k$ 。故有

$$\widehat{r}_k = \frac{\widehat{Ared}_k}{\widehat{Pred}_k} \geq \frac{Ared_k}{Pred_k} \rightarrow 1.$$

鉴于 μ_k 的更新规则, 存在一个正的常数 $\bar{m} > m$, 使得 $\mu_k < \bar{m}$ 对所有充分大的 k 都成立, 这与(3.10)相矛盾, 即当 S 为无限集时, 假设(3.8)不真。

情形 2. S 是有限的。在这种情形下, 存在一个索引 \hat{k} , 使得对所有的 $k > \hat{k}$, 有 $\widehat{r}_k < p_0 < p_1$ 。于是, 通过算法 1 的第 5 步, 对所有的 $k > \hat{k}$, 都有 $\mu_{k+1} = 4\mu_k$, 因此

$$\lim_{k \rightarrow \infty} \mu_k = +\infty. \quad (3.12)$$

由(3.4)和(3.8), 我们有

$$\|F_k\| \geq \frac{\|J_k^T F_k\|}{L} \geq \frac{\eta}{L}, \quad \forall k > \bar{k}.$$

从而

$$(1-\theta)\|F_k\|^\delta + \theta\|J_k^T F_k\|^\delta \geq \frac{(1-\theta)\eta^\delta}{L^\delta} + \theta\eta^\delta > 0, \quad \forall k > \bar{k}.$$

再由(2.11)和(3.12)可得

$$\lim_{k \rightarrow \infty} \lambda_k = +\infty.$$

上式结合(2.12)和(3.8)可得

$$\lim_{k \rightarrow \infty} d_k = 0.$$

类似于(3.11), 存在一个正的常数 \tilde{c} , 使得对所有充分大的 $k \in S$, 都有

$$\|\hat{d}_k\| \leq \tilde{c}\|d_k\|.$$

因此对所有充分大的 $k \in S$, 有

$$\|s_k\| = \|d_k + \hat{d}_k\| \leq (1+\tilde{c})\|d_k\|.$$

通过与情形 1 相同的分析, 我们可以得到 $\widehat{r}_k \rightarrow 1$ 。故存在一个正的常数 $\tilde{m} > m$, 使得 $\mu_k < \tilde{m}$ 对所有充分大的 k 都成立, 这与(3.12)相矛盾, 即当 S 为有限集时, 假设(3.8)不真。

总结情形 1 和情形 2, 我们得到(3.7)并完成了证明。 \square

4. 数值实验

为了验证算法的有效性, 我们将文献[11]中的单步 LM 算法 2.1 (记作算法 SLM)与本文提出的两步 LM 算法 1 (记作算法 TLM)进行数值实验比较。所有程序均在 MATLAB R2016a 平台上编写, 数值结果均在个人电脑上实现。算法所涉及的参数具体如下:

$$\mu_0 = 10^{-3}, m_0 = 10^{-8}, \tau = 0.5, p_0 = 10^{-4}, p_1 = 0.25, p_2 = 0.75.$$

我们选取算法终止条件为 $\|J_k^T F_k\| \leq 10^{-6}$, 计算以下两个具有不同初始点和不同大小的问题, 测试函数均选自[14]:

(i) 扩展的 Rosenbrock 函数: $F(x) = (F_1(x), \dots, F_n(x))^T$, 其中 n 是偶数, $i = 1, \dots, \frac{n}{2}$,

$$F_{2i-1}(x) = 10(x_{2i} - x_{2i-1}^2),$$

$$F_{2i}(x) = 1 - x_{2i-1}.$$

(ii) 扩展的 Powell Singular 函数: $F(x) = (F_1(x), \dots, F_n(x))^T$, 其中 n 是正整数的四倍, $i = 1, \dots, \frac{n}{4}$,

$$F_{4i-3}(x) = x_{4i-3} + 10x_{4i-2},$$

$$F_{4i-2}(x) = \sqrt{5}(x_{4i-1} - x_{4i}),$$

$$F_{4i-1}(x) = (x_{4i-2} - 2x_{4i-1})^2,$$

$$F_{4i}(x) = \sqrt{10}(x_{4i-3} - x_{4i})^2.$$

表 1~3 列出了两种算法在两个测试问题上的迭代次数, 测试 $\theta = 0, 0.5, 1$ 和 $\delta = 0.5, 1.0, 1.5, 2.0, 2.5$ 时的 LM 参数。我们选择 $x_0 = (-1, 1, -1, 1, \dots, -1, 1)$, 表 1~3 中的第三列表示不同的初始点, 对于问题(i), 取初始点为 $-10x_0, -x_0, 0x_0, x_0, 10x_0, 100x_0$; 对于问题(ii), 取初始点为 $x_0, 5x_0, 10x_0, 50x_0, 100x_0, 150x_0$ 。如果算法的迭代次数超过 1000, 则用符号 "--" 表示。

从表 1~3 中, 我们可以看到, 无论在测试问题(i)还是(ii)上, 算法 TLM 的迭代次数几乎都比算法 SLM 的迭代次数少, 说明本文提出的两步 LM 算法比单步 LM 算法的收敛速度更快。因此可以认为算法 1 在实际问题中可行且有效。

Table 1. Number of iterations at $\theta = 0$

表 1. $\theta = 0$ 时的迭代次数

Problem	n	x_0	$\delta = 0.5$	$\delta = 1.0$	$\delta = 1.5$	$\delta = 2.0$	$\delta = 2.5$
			SLM/TLM	SLM/TLM	SLM/TLM	SLM/TLM	SLM/TLM
Rosenbrock	2	-10	4/3	4/3	6/4	6/5	8/10
		-1	2/1	2/1	3/2	3/2	4/2
		0	15/1	17/1	17/1	17/1	17/1
		1	22/2	25/2	26/18	24/22	27/22
		10	4/3	5/3	6/4	7/4	7/5
		100	6/5	8/5	8/6	11/8	15/10
	10	-10	4/3	5/4	6/5	8/11	9/11
		-1	2/1	3/2	4/2	4/2	4/3
		0	17/1	16/1	16/1	18/1	20/1
		1	27/2	22/22	26/17	18/21	22/15
		10	5/3	5/4	6/5	8/7	9/7
		100	7/5	8/6	9/7	12/9	20/13
100	-10	4/3	6/4	7/6	8/6	11/12	

续表

		-1	2/1	3/2	4/3	5/3	6/3
		0	16/2	17/2	16/2	11/2	19/10
		1	27/19	22/22	24/18	23/14	21/16
		10	5/3	6/4	7/6	10/16	11/20
		100	7/5	8/6	10/10	16/14	28/25
		1	10/7	10/7	10/7	10/7	10/7
		5	12/9	12/9	12/9	12/9	13/9
	4	10	13/9	13/9	13/9	13/10	15/11
		50	15/11	15/11	15/11	17/12	20/15
		100	16/12	16/12	16/12	19/14	23/18
		150	17/12	17/12	17/12	20/14	24/19
		1	11/8	11/8	11/8	11/8	11/8
		5	13/9	13/9	13/9	13/10	16/12
Powell Singular	100	10	14/10	14/10	14/10	15/11	18/14
		50	16/12	16/12	16/12	20/15	24/19
		100	17/12	17/12	18/13	22/16	27/21
		150	18/13	18/13	18/13	23/17	32/24
		1	11/8	11/8	11/8	11/8	11/8
		5	13/9	13/9	13/9	14/10	16/12
	200	10	14/10	14/10	14/10	16/11	19/14
		50	16/12	16/12	17/12	20/15	25/20
		100	17/12	17/12	18/13	22/17	30/23
		150	18/13	18/13	19/13	23/18	39/28

Table 2. Number of iterations at $\theta = 0.5$

表 2. $\theta = 0.5$ 时的迭代次数

Problem	n	x_0	$\delta = 0.5$	$\delta = 1.0$	$\delta = 1.5$	$\delta = 2.0$	$\delta = 2.5$
			SLM/TLM	SLM/TLM	SLM/TLM	SLM/TLM	SLM/TLM
Rosenbrock	2	-10	4/3	6/5	8/7	11/9	16/16
		-1	2/1	4/2	4/3	5/3	7/4
		0	17/1	16/1	21/1	18/1	23/1
		1	18/2	23/2	27/22	26/23	30/17

续表

		10	5/3	6/5	8/5	11/8	34/13
		100	7/5	9/7	14/10	140/29	--/--
		-10	4/3	6/5	8/7	12/12	25/19
		-1	3/2	4/2	4/3	6/3	8/6
	10	0	19/1	15/2	16/2	18/2	29/2
		1	24/2	20/22	26/18	27/17	27/26
		10	5/3	6/5	9/7	12/10	24/16
		100	7/5	9/7	16/12	224/78	--/--
		-10	5/3	7/6	9/7	14/11	158/92
		-1	3/2	4/3	6/4	7/5	11/7
	100	0	16/2	14/2	16/3	20/3	12/12
		1	30/19	30/20	19/18	25/16	33/15
		10	5/4	7/5	9/8	16/15	108/59
		100	8/6	10/7	23/17	--/686	--/--
		1	10/7	10/7	10/7	10/7	11/8
		5	12/9	12/9	13/9	15/12	19/15
	4	10	13/9	13/9	14/10	18/14	24/19
		50	15/11	15/11	19/15	28/22	--/--
		100	16/12	17/12	22/17	87/52	--/--
		150	17/12	17/12	23/18	334/176	--/--
		1	11/8	11/8	11/8	12/9	14/11
		5	13/9	13/9	14/11	19/15	25/20
	100	10	14/10	14/10	17/12	22/17	148/82
		50	16/12	16/12	22/17	121/69	--/--
		100	17/12	18/13	24/19	--/777	--/--
		150	18/13	18/13	25/20	--/--	--/--
		1	11/8	11/8	11/8	12/9	15/12
		5	13/9	13/9	15/11	19/15	29/22
	200	10	14/10	14/10	17/13	22/18	318/168
		50	16/12	17/12	22/17	216/117	--/--
		100	17/12	18/13	25/19	--/--	--/--
		150	18/13	19/13	26/21	--/--	--/--

Table 3. Number of iterations at $\theta = 1$
表 3. $\theta = 1$ 时的迭代次数

Problem	n	x_0	$\delta = 0.5$	$\delta = 1.0$	$\delta = 1.5$	$\delta = 2.0$	$\delta = 2.5$
			SLM/TLM	SLM/TLM	SLM/TLM	SLM/TLM	SLM/TLM
Rosenbrock	2	-10	4/3	6/5	9/6	11/10	20/17
		-1	2/2	4/2	4/3	6/4	7/4
		0	16/1	18/1	18/1	22/1	33/1
		1	22/2	26/2	21/22	39/17	36/26
		10	5/3	6/5	8/6	13/9	18/15
		100	7/5	9/7	16/12	109/41	--/--
	10	-10	4/3	6/5	9/8	12/11	37/27
		-1	3/2	4/2	5/3	6/4	8/6
		0	15/2	19/2	23/2	23/2	19/2
		1	25/2	29/17	27/18	27/15	39/23
		10	5/4	7/5	9/7	13/10	29/19
		100	8/6	10/7	18/14	382/166	--/--
	100	-10	5/4	7/6	11/9	18/14	295/156
		-1	3/2	4/3	7/4	9/6	11/8
		0	16/2	20/2	16/3	16/3	25/11
		1	25/21	29/18	26/21	23/19	38/22
		10	6/4	7/6	10/8	16/13	192/100
		100	8/6	10/8	20/77	--/--	--/--
Powell Singular	4	1	10/7	10/7	10/7	10/7	11/8
		5	12/9	12/9	13/9	16/12	20/16
		10	13/9	13/9	15/11	19/15	27/21
		50	15/11	16/11	20/15	33/24	--/--
		100	16/12	17/12	22/17	148/83	--/--
	100	150	17/12	17/12	23/18	639/329	--/--
		1	11/8	11/8	11/8	12/9	15/11
		5	13/9	13/9	15/11	19/15	28/21
		10	14/10	14/10	17/13	22/18	272/144
		50	16/12	17/12	22/17	216/117	--/--

续表

	100	17/12	18/13	25/19	--/--	--/--
	150	18/13	18/13	26/21	--/--	--/--
200	1	11/8	11/8	11/8	13/10	16/12
	5	13/9	13/9	15/11	20/16	36/26
	10	14/10	14/10	18/13	23/18	612/315
	50	16/12	17/12	23/18	405/211	--/--
	100	17/12	18/13	25/20	--/--	--/--
	150	18/13	19/14	27/21	--/--	--/--

5. 总结

本文通过采用一种较为通用的 LM 参数并结合非单调信赖域方法, 提出了改进型的两步 Levenberg-Marquardt 方法。我们已经证明, 新的两步 LM 方法具有全局收敛性, 并通过与单步 LM 方法的数值实验进行对比, 得出新的两步 LM 方法在解决实际问题上高效且有前途。

参考文献

- [1] Ortega, J.M. and Rheinboldt, W.C. (2000) Iterative Solution of Nonlinear Equations in Several Variables. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9780898719468>
- [2] Kelley, C.T. (2003) Solving Nonlinear Equations with Newton's Method. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9780898718898>
- [3] Yamashita, N. and Fukushima, M. (2001) On the Rate of Convergence of the Levenberg-Marquardt Method. In: Alefeld, G. and Chen, X., Eds., *Topics in Numerical Analysis*, Springer, 239-249. https://doi.org/10.1007/978-3-7091-6217-0_18
- [4] Fan, J. and Yuan, Y. (2005) On the Quadratic Convergence of the Levenberg-Marquardt Method without Nonsingularity Assumption. *Computing*, **74**, 23-39. <https://doi.org/10.1007/s00607-004-0083-1>
- [5] Ma, C. and Jiang, L. (2007) Some Research on Levenberg-Marquardt Method for the Nonlinear Equations. *Applied Mathematics and Computation*, **184**, 1032-1040. <https://doi.org/10.1016/j.amc.2006.07.004>
- [6] Fan, J. and Pan, J. (2009) A Note on the Levenberg-Marquardt Parameter. *Applied Mathematics and Computation*, **207**, 351-359. <https://doi.org/10.1016/j.amc.2008.10.056>
- [7] Fan, J. (2012) The Modified Levenberg-Marquardt Method for Nonlinear Equations with Cubic Convergence. *Mathematics of Computation*, **81**, 447-466. <https://doi.org/10.1090/s0025-5718-2011-02496-8>
- [8] Grippo, L., Lampariello, F. and Lucidi, S. (1986) A Nonmonotone Line Search Technique for Newton's Method. *SIAM Journal on Numerical Analysis*, **23**, 707-716. <https://doi.org/10.1137/0723046>
- [9] Deng, N.Y., Xiao, Y. and Zhou, F.J. (1993) Nonmonotonic Trust Region Algorithm. *Journal of Optimization Theory and Applications*, **76**, 259-285. <https://doi.org/10.1007/bf00939608>
- [10] Zhang, H. and Hager, W.W. (2004) A Nonmonotone Line Search Technique and Its Application to Unconstrained Optimization. *SIAM Journal on Optimization*, **14**, 1043-1056. <https://doi.org/10.1137/s1052623403428208>
- [11] Zhao, L. and Tang, J. (2024) Levenberg-Marquardt Method with a General LM Parameter and a Nonmonotone Trust Region Technique. *Journal of Applied Analysis & Computation*, **14**, 1959-1976. <https://doi.org/10.11948/20220441>
- [12] Fan, J. (2003) A Modified Levenberg-Marquardt Algorithm for Singular System of Nonlinear Equations. *Journal of Computational Mathematics*, **21**, 625-636.
- [13] Powell, M.J.D. (1975) Convergence Properties of a Class of Minimization Algorithms. In: Mangasarian, O.L., Meyer, R.R. and Robinson, S.M., Eds., *Nonlinear Programming 2*, Academic Press, 1-27. <https://doi.org/10.1016/B978-0-12-468650-2.50005-5>
- [14] Moré, J.J., Garbow, B.S. and Hillstom, K.E. (1981) Testing Unconstrained Optimization Software. *ACM Transactions on Mathematical Software*, **7**, 17-41. <https://doi.org/10.1145/355934.355936>