doi：10.13413/j.cnki.jdxblxb.2023338

# 求解非线性方程组的非单调自适应加速 Levenberg-Marquardt 算法

曹名圆[1]，李 蓉[1]，闫雪丽[1]，黄庆道[2]
（1. 北华大学 数学与统计学院，吉林 吉林 132013；2. 吉林大学 数学学院，长春 130012）

**摘要**：提出一种新的求解非线性方程组的非单调自适应加速 Levenberg-Marquardt 算法，该算法使用一种新的自适应函数更新 Levenberg-Marquardt 参数，这种 Levenberg-Marquardt 参数的更新方式可提高过于成功的迭代中模型与目标函数的一致性，从而加快算法的收敛速度. 数值实验结果表明，该算法具有良好的数值计算性能.

**关键词**：自适应函数；非单调技术；加速 Levenberg-Marquardt 算法
**中图分类号**：O221.2 **文献标志码**：A **文章编号**：1671-5489(2024)03-0538-09

# Nonmonotonic Adaptive Accelerated Levenberg-Marquardt Algorithm for Solving Nonlinear Equations

CAO Mingyuan[1], LI Rong[1], YAN Xueli[1], HUANG Qingdao[2]
(1. *School of Mathematics and Statistics, Beihua University, Jilin* 132013, *Jilin Province, China*;
2. *College of Mathematics, Jilin University, Changchun* 130012, *China*)

**Abstract**: We proposed a new nonmonotonic adaptive accelerated Levenberg-Marquardt algorithm for solving nonlinear equations. The algorithm used a new adaptive function to update the Levenberg-Marquardt parameter, which could enhance the consistency between the model and objective function during too-successful iterations, thereby accelerating the convergence rate of the algorithm. Numerical experimental results show that the proposed algorithm has good numerical computational performance.

**Keywords**: adaptive function; nonmonotonic technique; accelerated Levenberg-Marquardt algorithm

## 0 引 言

考虑非线性方程组

$$\boldsymbol{F}(\boldsymbol{x})=0, \tag{1}$$

其中 $\boldsymbol{F}(\boldsymbol{x})$：$\mathbb{R}^n\to\mathbb{R}^m$ 是一个连续可微函数. Levenberg-Marquardt(LM)方法[1-2]是求解方程组(1)常用的迭代方法，该方法在第 $k$ 次迭代中计算试探步为

$$\boldsymbol{d}_k=-(\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{J}_k+\lambda_k\boldsymbol{I})^{-1}(\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k), \tag{2}$$

**收稿日期**：2023-08-20.

**第一作者简介**：曹名圆(1986—)，女，满族，博士，副教授，从事最优化理论及其应用的研究，E-mail：cmy0918@beihua.edu.cn.
**通信作者简介**：黄庆道(1968—)，男，汉族，博士，教授，从事最优化理论及其应用的研究，E-mail：huangqd@jlu.edu.cn.

**基金项目**：吉林省自然科学基金重点项目(批准号：YDZJ202101ZYTS167；YDZJ202201ZYTS303)和北华大学研究生创新项目(批准号：2022033).

其中 $\boldsymbol{F}_k=\boldsymbol{F}(\boldsymbol{x}_k)$，$\boldsymbol{J}_k=\boldsymbol{J}(\boldsymbol{x}_k)$是 $\boldsymbol{F}(\boldsymbol{x})$在 $\boldsymbol{x}_k$ 处的 Jacobi 矩阵，$\boldsymbol{I}$ 是一个单位矩阵，$\lambda_k>0$ 是 LM 参数. LM 参数的选择对 LM 方法的数值性能至关重要，目前已经有许多有效的 LM 参数[3-4].

为克服单调 LM 算法的局限性，许多研究者将非单调技术引入到 LM 方法中. Chamberlain 等[5]首先提出约束优化的 Watchdog 技术，放宽了标准线搜索条件，以克服约束优化的 marotos 效应；Grippo 等[6]提出了应用于牛顿方法的非单调线搜索技术，并将该技术应用于截断牛顿方法[7]中；Amini 等[8]提出了针对 LM 方法的非单调线搜索技术；Kimiaei[9]对文献[8]中的非单调技术进行了改进，将新的非单调技术应用于 LM 方法中；Huang 等[10]将提出的非单调的 LM 方法与文献[11]中的单调 LM 方法进行了比较，数值实验结果表明，带有非单调技术的算法比没有非单调技术的算法更有效. 因此，本文将在 LM 方法中引入非单调技术，在保证全局收敛性的前提下减少总迭代次数和运行时间.

除上述非单调技术的引入外，自适应技术的引入也能减少算法迭代次数，加快算法的收敛速度. Hei[12]在信赖域方法中提出了一个 $R$ 函数，使用自适应更新技术更新信赖域半径 $\Delta_k$，即 $\Delta_{k+1}=R(r_k)\Delta_k$. 此外，Walmag 等[13]提出了一个 $\Lambda$ 函数更新信赖域半径，即 $\Delta_{k+1}=\Lambda(r_k)\Delta_k$，其中 $\Lambda$ 是一个关于 $r_k$ 的非负有界函数. 在此基础上，Lu 等[14]研究表明，在过于成功的迭代中，模型与目标函数之间的一致性欠佳，因此提出了一个 $L$ 函数更新信赖域半径. 结果表明，$L$ 函数包含了 $R$ 函数和 $\Lambda$ 函数的一些有利特征，并且该方法在过于成功的迭代中更有效.

本文将构造一个新的自适应更新函数更新 LM 参数以提高算法的效率. 首先将一种有效的非单调技术应用于 LM 方法，该方法在不影响算法全局收敛性的情况下允许函数值有所增长；然后选择一种有效的 LM 参数更新试探步，当迭代序列远离最优解集时，接受大量次不成功的迭代，避免迭代点在局部区域来回跳跃的情况；最后构造一种自适应函数更新参数 $\mu_k$，充分利用实际下降量与预测下降量的比值信息判断是否接受试探步，提高了过于成功的迭代中模型与目标函数的一致性. 此外，证明了本文的非单调自适应加速 LM 算法在适当条件下具有全局收敛性，并在局部误差界条件下保持二阶收敛速率. 数值计算结果表明，该方法具有较好的数值性能.

## 1　算法设计

定义价值函数 $\|\boldsymbol{F}_k\|^2$ 在第 $k$ 次迭代的实际下降量与预测下降量之间的比值为 $r_k=\dfrac{\mathrm{Ared}_k}{\mathrm{Pred}_k}$，该比值决定了是否接受试探步 $\boldsymbol{d}_k$，其中实际下降量和预测下降量分别为

$$\mathrm{Ared}_k=\|\boldsymbol{F}_k\|^2-\|\boldsymbol{F}(\boldsymbol{x}_k+\boldsymbol{d}_k)\|^2, \tag{3}$$

$$\mathrm{Pred}_k=\|\boldsymbol{F}_k\|^2-\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|^2. \tag{4}$$

本文将非单调技术应用于 LM 方法，使用实际下降量

$$\overline{\mathrm{Ared}_k}=\Lambda_k-\|\boldsymbol{F}(\boldsymbol{x}_k+\boldsymbol{d}_k)\|^2, \tag{5}$$

这里

$$\Lambda_k=\begin{cases}\|\boldsymbol{F}_k\|^2, & k=0,\\ \dfrac{\sum\limits_{i=0}^{m(k)-1}\eta^{m(k)-i}\mathscr{F}_k(i)+\|\boldsymbol{F}_k\|^2}{\sum\limits_{i=0}^{m(k)-1}\eta^{m(k)-i}+1}, & k>0,\end{cases} \tag{6}$$

其中：$m(0)=0$，$0\leqslant m(k)\leqslant\min\{m(k-1)+1,N\}$，$N\in\mathbb{N}$；$\eta\in[\eta_{\min},\eta_{\max}]$，$\eta_{\min}\in[0,1)$，$\eta_{\max}\in[\eta_{\min},1]$；

$$\mathscr{F}_k=\{\|\boldsymbol{F}_{k-j}\|^2\}_{0\leqslant j\leqslant m(k)},\quad k\in\mathbb{N};\qquad \mathscr{F}_k(i)=\begin{cases}\|\boldsymbol{F}_k\|^2, & k<N,\\ \|\boldsymbol{F}_{k-N+i+1}\|^2, & k\geqslant N,\end{cases}$$

式中

$$i\in\begin{cases}[0,k], & k<N,\\ [0,N-1], & k\geqslant N.\end{cases}$$

为存储新的信息并去除旧信息，非单调技术利用序列$\{\mathscr{F}_k\}$确定阻尼参数. 本文所用的比值为

$$\bar{r}_k=\frac{\overline{\mathrm{Ared}_k}}{\mathrm{Pred}_k}=\frac{\Lambda_k-\|\boldsymbol{F}(\boldsymbol{x}_k+\boldsymbol{d}_k)\|^2}{\|\boldsymbol{F}_k\|^2-\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|^2}. \tag{7}$$

在LM方法中使用非单调线搜索技术，可以在连续迭代时不强制执行目标函数值的严格单调性，即函数值不需要在每次迭代中严格减少，并且允许较大的步长，因此它可能是整个优化过程的一个更好选择. 特别是当所求问题存在弯曲狭谷的情况下，非单调技术可增加算法寻找到全局最优值的可能性并提高算法的收敛速度.

本文选择LM参数

$$\lambda_{k+1}=\frac{\mu_{k+1}\|\boldsymbol{F}_{k+1}\|^2}{1+\|\boldsymbol{F}_{k+1}\|^2}. \tag{8}$$

显然，当$\{\boldsymbol{x}_k\}$远离解集，$\|\boldsymbol{F}_k\|$非常大时，$\frac{\|\boldsymbol{F}_{k+1}\|^2}{1+\|\boldsymbol{F}_{k+1}\|^2}$接近于1，$\lambda_k$接近于$\mu_k$，选择这个LM参数能提高算法的效率.

在过于成功的迭代中，目标函数的实际下降量明显大于预测下降量. 虽然目前的迭代允许算法向最优方向发展，但模型函数对目标函数的近似并不准确. 所以为克服迭代在过于成功时的缺陷，本文构造了更新参数$\mu_{k+1}$的自适应函数：

$$K(\bar{r}_k)=\begin{cases}\beta_1+(\beta_2-\beta_1)\exp\left\{-\left(\dfrac{-\bar{r}_k+p_1}{p_1}\right)^2\right\}, & \bar{r}_k\leqslant p_1,\\ \beta_2, & p_1<\bar{r}_k<p_2,\\ \dfrac{1-\beta_3\exp\{p_2\}}{1-\exp\{p_2\}}-\dfrac{(1-\beta_3)\exp\{p_2\}}{1-\exp\{p_2\}}\exp\{-\bar{r}_k+p_2\}-\dfrac{1}{2}, & \bar{r}_k\geqslant p_2,\end{cases} \tag{9}$$

其中$\beta_1,\beta_2,\beta_3,p_1,p_2$是常数，并且$0<\beta_2<1<\beta_1\leqslant\beta_3$，$0<p_1<p_2<1$. 因此，$K(\bar{r}_k)$满足以下性质：

1) $\lim\limits_{\bar{r}_k\to\infty}K(\bar{r}_k)=\beta_1$；2) $\lim\limits_{\bar{r}_k\to p_1}K(\bar{r}_k)=\beta_2$；3) $\lim\limits_{\bar{r}_k\to p_2}K(\bar{r}_k)=\dfrac{1}{2}$；4) $\lim\limits_{\bar{r}_k\to+\infty}K(\bar{r}_k)=\dfrac{1-\beta_3\exp\{p_2\}}{1-\exp\{p_2\}}-\dfrac{1}{2}$.

本文参数$\mu_{k+1}$的更新方式为

$$\mu_{k+1}=\max\{m,K(\bar{r}_k)\mu_k\}. \tag{10}$$

当比值$\bar{r}_k$接近于1，即模型函数提供了目标函数的精确局部近似时，接受试探步$\boldsymbol{d}_k$并减小$\mu_k$；否则，拒绝试探步$\boldsymbol{d}_k$并增大$\mu_k$. 在这种参数$\mu_{k+1}$的更新方式中，当比值$\bar{r}_k$充分大于1，即迭代过于成功时，选择增大$\mu_k$可提高在过于成功的迭代中模型与目标函数的一致性，避免迭代点在局部跳跃，加快算法的收敛速度，增强算法的有效性和稳定性.

由式(2)可知试探步$\boldsymbol{d}_k$是凸优化问题：

$$\min_{\boldsymbol{d}\in\mathbb{R}^n}\{\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}\|^2+\lambda_k\|\boldsymbol{d}\|^2\}\doteq\varphi_k(\boldsymbol{d})$$

的解，如果$\Delta_k=\|-(\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{J}_k+\lambda_k\boldsymbol{I})^{-1}(\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k)\|$，则$\boldsymbol{d}_k$也是信赖域子问题

$$\min_{\boldsymbol{d}\in\mathbb{R}^n}\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}\|^2,$$
$$\text{s.t.}\quad\|\boldsymbol{d}\|\leqslant\Delta_k$$

的解. 根据文献[15]的结论，有

$$\mathrm{Pred}_k\geqslant\|\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k\|\min\left\{\|\boldsymbol{d}_k\|,\frac{\|\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k\|}{\|\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{J}_k\|}\right\}. \tag{11}$$

在上述分析的基础上，本文提出一种新的非单调自适应加速LM(nonmonotone adaptive accelerated Levenberg-Marquardt，NALM)算法.

**算法1** NALM算法.

初始点选取：给定$\boldsymbol{x}_0\in\mathbb{R}^n$，$\mu_0>m>0$，$0\leqslant p_0<p_1<p_2<1$，$0<\beta_2<1<\beta_1\leqslant\beta_3$，$\varepsilon>0$. 令$k:=0$.

步骤1) 计算$\boldsymbol{F}_k$和$\boldsymbol{J}_k$. 如果$\|\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k\|\leqslant\varepsilon$，则停止计算；否则，通过式(8)求自适应LM参数$\lambda_k$.

步骤2) 求解

$$(\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{J}_k+\lambda_k\boldsymbol{I})\boldsymbol{d}=-\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k \tag{12}$$

得到 $\boldsymbol{d}_k$.

步骤 3) 利用式(4),(5),(7)分别计算 $\mathrm{Pred}_k$,$\overline{\mathrm{Ared}_k}$和 $\bar{r}_k$.

步骤 4) 设

$$\boldsymbol{x}_{k+1}=\begin{cases}\boldsymbol{x}_k+\boldsymbol{d}_k, & \bar{r}_k\geqslant p_0,\\ \boldsymbol{x}_k, & \bar{r}_k<p_0.\end{cases}\tag{13}$$

步骤 5) 利用式(9)计算 $K(\bar{r}_k)$，并利用式(10)更新参数 $\mu_{k+1}$.

令 $k:=k+1$，返回步骤 1).

## 2　收敛性分析

### 2.1　全局收敛性

**假设 1**　(i) 对任意给定的 $\boldsymbol{x}_0\in\mathbb{R}^n$，水平集 $L(\boldsymbol{x}_0):=\{\boldsymbol{x}\in\mathbb{R}^n\mid\|\boldsymbol{F}(\boldsymbol{x})\|\leqslant\|\boldsymbol{F}(\boldsymbol{x}_0)\|\}$ 是有界的.

(ii) $\boldsymbol{F}(\boldsymbol{x})$是连续可微函数，$\boldsymbol{F}(\boldsymbol{x})$和 Jacobi 矩阵 $\boldsymbol{J}(\boldsymbol{x})$是 Lipschitz 连续的，即存在正常数 $L_1$ 和 $L_2$，满足

$$\|\boldsymbol{J}(\boldsymbol{x})-\boldsymbol{J}(\boldsymbol{y})\|\leqslant L_1\|\boldsymbol{x}-\boldsymbol{y}\|,\qquad\forall\,\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n,\tag{14}$$

$$\|\boldsymbol{F}(\boldsymbol{x})-\boldsymbol{F}(\boldsymbol{y})\|\leqslant L_2\|\boldsymbol{x}-\boldsymbol{y}\|,\qquad\forall\,\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n.\tag{15}$$

由式(14),(15)，可得

$$\|\boldsymbol{F}(\boldsymbol{y})-\boldsymbol{F}(\boldsymbol{x})-\boldsymbol{J}(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x})\|\leqslant L_1\|\boldsymbol{y}-\boldsymbol{x}\|^2,\qquad\forall\,\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n,\tag{16}$$

$$\|\boldsymbol{J}(\boldsymbol{x})\|\leqslant L_2,\qquad\forall\,\boldsymbol{x}\in\mathbb{R}^n.\tag{17}$$

记 $NF_{l(k)}:=\max\{\mathscr{F}_k\}$，$k\in\mathbb{N}$.

**引理 1**　若序列$\{\boldsymbol{x}_k\}$由 NALM 算法产生，则：

1) 对所有的 $k\in\mathbb{N}$，有 $\boldsymbol{x}_k\in L(\boldsymbol{x}_0)$；

2) $\{NF_{l(k)}\}_{k\geqslant0}$是一个收敛序列.

证明：1) 首先证明

$$\Lambda_k\leqslant NF_{l(k)}.\tag{18}$$

由定义，有

$$\Lambda_k=\frac{\sum\limits_{i=0}^{m(k)-1}\eta^{m(k)-i}\mathscr{F}_k(i)+\|\boldsymbol{F}_k\|^2}{\sum\limits_{i=0}^{m(k-1)}\eta^{m(k)-i}+1}\leqslant\frac{\sum\limits_{i=0}^{m(k)-1}\eta^{m(k)-i}NF_{l(k)}+NF_{l(k)}}{\sum\limits_{i=0}^{m(k-1)}\eta^{m(k)-i}+1}=NF_{l(k)},$$

即式(18)成立. 另一方面，由于 $\boldsymbol{x}_{k+1}$被算法接受，故

$$\Lambda_k-\|\boldsymbol{F}_{k+1}\|^2\geqslant\|\boldsymbol{F}_k\|^2-\|\boldsymbol{F}_{k+1}\|^2=\mathrm{Ared}_k\geqslant p_0\mathrm{Pred}_k>0,$$

即

$$\|\boldsymbol{F}_{k+1}\|^2\leqslant\Lambda_k.\tag{19}$$

下面使用数学归纳法证明. 当 $k=0$ 时，有 $\Lambda_0=\|\boldsymbol{F}_0\|^2$，所以 $\|\boldsymbol{F}_1\|^2\leqslant\|\boldsymbol{F}_0\|^2$ 成立. 假设对所有的$i=1,2,\cdots,k$，都有 $\boldsymbol{x}_k\in L(\boldsymbol{x}_0)$，即对所有的 $i=1,2,\cdots,k$，都有 $\|\boldsymbol{F}_i\|^2\leqslant\|\boldsymbol{F}_0\|^2$ 成立，则 $NF_{l(k)}\leqslant\|\boldsymbol{F}_0\|^2$. 结合式(18),(19)，可得

$$\|\boldsymbol{F}_{k+1}\|^2\leqslant\Lambda_k\leqslant NF_{l(k)}\leqslant\|\boldsymbol{F}_0\|^2.$$

因此 $\boldsymbol{x}_{k+1}\in L(\boldsymbol{x}_0)$，所以有$\{\boldsymbol{x}_k\}_{k\geqslant0}$包含在 $L(\boldsymbol{x}_0)$中.

2) 由式(18),(19)，有

$$NF_{l(k+1)}=\max\{\mathscr{F}_{k+1}\}=\max_{0\leqslant j\leqslant N}\{\|\boldsymbol{F}_{k+1-j}\|^2\}\leqslant\max\{\max_{0\leqslant j\leqslant N}\{\|\boldsymbol{F}_{k-j}\|^2\},\|\boldsymbol{F}_{k+1}\|^2\}=$$

$$\max\{NF_{l(k)},\Lambda_k\}=NF_{l(k)},$$

所以$\{NF_{l(k)}\}_{k\geqslant0}$是一个递减序列，由假设 1 中(i)和 $\boldsymbol{x}_k\in L(\boldsymbol{x}_0)$可得$\{NF_{l(k)}\}_{k\geqslant0}$是一个收敛序列.

**定理 1**　在假设 1 的条件下，由 NALM 算法生成序列$\{\boldsymbol{x}_k\}$，则

$$\lim_{k\to\infty}\| \boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k \| =0. \tag{20}$$

证明：反证法. 假设存在一个正常数 $\varepsilon>0$ 和无穷多个 $k$，使得

$$\| \boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k \| \geqslant \varepsilon. \tag{21}$$

首先证明

$$\lim_{k\to\infty} \Lambda_k =\lim_{k\to\infty}\| \boldsymbol{F}_k \|^2. \tag{22}$$

由引理1中2)以及 $\Lambda_k\leqslant NF_{l(k)}$ 可知 $\Lambda_k$ 存在极限. 对不等式(18),(19)两边取极限可得

$$\lim_{k\to\infty}\| \boldsymbol{F}_k \|^2 =\lim_{k\to\infty}\| \boldsymbol{F}_{k+1} \|^2 \leqslant \lim_{k\to\infty}\Lambda_k, \qquad \lim_{k\to\infty}\Lambda_k \leqslant \lim_{k\to\infty} NF_{l(k)} \leqslant \lim_{k\to\infty}\| \boldsymbol{F}_k \|^2.$$

结合上述两个不等式可知式(22)成立.

由于 $\boldsymbol{d}_k$ 被算法接受，所以 $\Lambda_k-\| \boldsymbol{F}_{k+1} \|^2\geqslant p_0\mathrm{Pred}_k$. 根据式(11),(17),(21)，可知

$$\Lambda_k-\| \boldsymbol{F}_{k+1} \|^2 \geqslant p_0\| \boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k \| \min\left\{\| \boldsymbol{d}_k \|, \frac{\| \boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k \|}{\| \boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{J}_k \|}\right\} \geqslant p_0\varepsilon\min\left\{\| \boldsymbol{d}_k \|, \frac{\varepsilon}{L_2^2}\right\}. \tag{23}$$

结合不等式(23),(22)，可得

$$\lim_{k\to\infty}\boldsymbol{d}_k =\boldsymbol{0}. \tag{24}$$

从而由式(2),(8),(17),(21)，有

$$\lim_{k\to\infty}\mu_k =+\infty. \tag{25}$$

另一方面，由式(3),(11),(17),(21),(24)，有

$$|r_k-1|=\left|\frac{\mathrm{Ared}_k}{\mathrm{Pred}_k}-1\right|=\left|\frac{\| \boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k \|^2-\| \boldsymbol{F}_{k+1} \|^2}{\mathrm{Pred}_k}\right|=$$

$$\frac{\| \boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k \| O(\| \boldsymbol{d}_k \|^2)+O(\| \boldsymbol{d}_k \|^4)}{\mathrm{Pred}_k}\leqslant$$

$$\frac{\| \boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k \| O(\| \boldsymbol{d}_k \|^2)+O(\| \boldsymbol{d}_k \|^4)}{\| \boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k \| \min\left\{\| \boldsymbol{d}_k \|, \frac{\| \boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k \|}{\| \boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{J}_k \|}\right\}}\leqslant\frac{O(\| \boldsymbol{d}_k \|^2)}{\| \boldsymbol{d}_k \|}\to 0.$$

再结合式(5)～(7)，可得

$$\bar{r}_k =\frac{\overline{\mathrm{Ared}_k}}{\mathrm{Pred}_k}=\frac{\Lambda_k-\| \boldsymbol{F}_{k+1} \|^2}{\mathrm{Pred}_k}\geqslant\frac{\| \boldsymbol{F}_k \|^2-\| \boldsymbol{F}_{k+1} \|^2}{\mathrm{Pred}_k}=r_k\to 1.$$

根据 $\mu_k$ 的更新规则，可知存在一个正常数 $\bar{\mu}>m$，使得对所有足够大的 $k$，$\mu_k<\bar{\mu}$ 都成立，这与式(25)矛盾，从而式(20)成立.

### 2.2 局部收敛性

下面利用奇异值分解技术研究算法的局部收敛性. 假设由算法生成的序列 $\boldsymbol{x}_k$ 收敛于非空解集 $X^*$，并且位于 $\boldsymbol{x}^*\in X^*$ 的某个邻域内.

**假设2** (i) $\boldsymbol{F}(\boldsymbol{x})$是连续可微函数，$\boldsymbol{J}(\boldsymbol{x})$在 $N(\boldsymbol{x}^*,b_1)$上 Lipschitz 连续，$b_1<1$，即存在一个正常数 $L_1$，满足

$$\| \boldsymbol{J}(\boldsymbol{y})-\boldsymbol{J}(\boldsymbol{x}) \| \leqslant L_1\| \boldsymbol{y}-\boldsymbol{x} \|, \qquad \forall\, \boldsymbol{x},\boldsymbol{y}\in N(\boldsymbol{x}^*,b_1)=\{\boldsymbol{x}\mid \| \boldsymbol{x}-\boldsymbol{x}^* \| \leqslant b_1\}. \tag{26}$$

(ii) $\| \boldsymbol{F}(\boldsymbol{x}) \|$ 在 $\boldsymbol{x}^*\in X^*$ 的某邻域内提供了一个方程组(1)的局部误差界，即存在一个正常数 $c_1>0$，满足

$$\| \boldsymbol{F}(\boldsymbol{x}) \| \geqslant c_1\mathrm{dist}(\boldsymbol{x},X^*), \qquad \forall\, \boldsymbol{x}\in N(\boldsymbol{x}^*,b_1), \tag{27}$$

其中 $\mathrm{dist}(\boldsymbol{x},X^*)$是 $\boldsymbol{x}$ 到 $X^*$ 的距离.

根据式(26)，有

$$\| \boldsymbol{F}(\boldsymbol{y})-\boldsymbol{F}(\boldsymbol{x})-\boldsymbol{J}(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}) \| \leqslant L_1\| \boldsymbol{y}-\boldsymbol{x} \|^2, \tag{28}$$

$$\| \boldsymbol{F}(\boldsymbol{y})-\boldsymbol{F}(\boldsymbol{x}) \| \leqslant L_2\| \boldsymbol{y}-\boldsymbol{x} \|, \qquad \forall\, \boldsymbol{x},\boldsymbol{y}\in N(\boldsymbol{x}^*,b_1), \tag{29}$$

其中 $L_2$ 是一个正常数. 下面使用 $\bar{\boldsymbol{x}}_k\in X^*$ 表示满足

$$\| \bar{\boldsymbol{x}}_k-\boldsymbol{x}_k \| =\mathrm{dist}(\boldsymbol{x}_k,X^*) \tag{30}$$

的向量. 为得到序列$\{\boldsymbol{x}_k\}$的局部收敛速度，提出以下引理.

**引理 2**　在假设 2 的条件下，对所有足够大的 $k$，存在一个常数 $c_2>0$，满足 $\|\boldsymbol{d}_k\|\leqslant c_2\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|$.

证明：由 $\bar{\boldsymbol{x}}_k$ 的定义式(30)可知

$$\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}^*\|\leqslant\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|+\|\boldsymbol{x}_k-\boldsymbol{x}^*\|\leqslant 2\|\boldsymbol{x}_k-\boldsymbol{x}^*\|\leqslant b_1,$$

即 $\bar{\boldsymbol{x}}\in N(\boldsymbol{x}^*,b_1)$. 根据式(13)，有

$$\lambda_k=\frac{\mu_k\|\boldsymbol{F}_k\|^2}{1+\|\boldsymbol{F}_k\|^2}=\mu_k\left(1-\frac{1}{1+\|\boldsymbol{F}_k\|^2}\right)\geqslant m\left(1-\frac{1}{1+c_1^2\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2}\right)=\frac{mc_1^2\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2}{1+c_1^2\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2}.$$

根据不等式(28)，得

$$\|\boldsymbol{F}_k+\boldsymbol{J}_k(\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k)\|^2=\|\boldsymbol{F}(\bar{\boldsymbol{x}}_k)-\boldsymbol{F}_k-\boldsymbol{J}_k(\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k)\|^2\leqslant L_1^2\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^4.$$

由于 $\boldsymbol{d}_k$ 是 $\varphi_k(\boldsymbol{d})$的最小点，因此

$$\|\boldsymbol{d}_k\|^2\leqslant\frac{1}{\lambda_k}\varphi_k(\boldsymbol{d}_k)\leqslant\frac{1}{\lambda_k}\varphi_k(\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k)=\frac{1}{\lambda_k}(\|\boldsymbol{F}_k+\boldsymbol{J}_k(\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k)\|^2+\lambda_k\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2)\leqslant$$

$$\frac{1+c_1^2\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2}{mc_1^2\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2}(L_1^2\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^4)+\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2=O(\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2).$$

所以存在一个常数 $c_2>0$，使得 $\|\boldsymbol{d}_k\|\leqslant c_2\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|$.

**引理 3**　在假设 2 的条件下，对所有足够大的 $k$，存在一个正常数 $M>m$，满足 $\mu_k\leqslant M$.

证明：首先证明对足够大的 $k$，下面的不等式成立：

$$\mathrm{Pred}_k=\|\boldsymbol{F}_k\|^2-\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|^2\geqslant\min\left\{\frac{c_1}{2c_2},\frac{c_1}{2}\right\}\|\boldsymbol{F}_k\|\,\|\boldsymbol{d}_k\|.\tag{31}$$

考虑如下两种情形.

情形 1) 如果 $\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|\leqslant\|\boldsymbol{d}_k\|$，则根据 $\boldsymbol{d}_k$ 的定义和假设 2，有

$$\|\boldsymbol{F}_k\|-\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|\geqslant\|\boldsymbol{F}_k\|-\|\boldsymbol{F}_k+\boldsymbol{J}_k(\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k)\|\geqslant$$

$$c_1\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|-L_1\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2\geqslant\frac{c_1}{2c_2}\|\boldsymbol{d}_k\|.\tag{32}$$

情形 2) 如果 $\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|>\|\boldsymbol{d}_k\|$，则有

$$\|\boldsymbol{F}_k\|-\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|\geqslant\|\boldsymbol{F}_k\|-\left\|\boldsymbol{F}_k+\frac{\|\boldsymbol{d}_k\|}{\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|}\boldsymbol{J}_k(\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k)\right\|\geqslant$$

$$\frac{\|\boldsymbol{d}_k\|}{\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|}(\|\boldsymbol{F}_k\|-\|\boldsymbol{F}_k+\boldsymbol{J}_k(\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k)\|)\geqslant$$

$$\frac{\|\boldsymbol{d}_k\|}{\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|}(c_1\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|-L_1\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2)\geqslant\frac{c_1}{2}\|\boldsymbol{d}_k\|.\tag{33}$$

结合不等式(32)，(33)以及引理 2 可知

$$\mathrm{Pred}_k=(\|\boldsymbol{F}_k\|+\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|)(\|\boldsymbol{F}_k\|-\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|)\geqslant$$

$$\|\boldsymbol{F}_k\|(\|\boldsymbol{F}_k\|-\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|)\geqslant\min\left\{\frac{c_1}{2c_2},\frac{c_1}{2}\right\}\|\boldsymbol{F}_k\|\,\|\boldsymbol{d}_k\|,\tag{34}$$

即式(31)成立. 因此，由式(31)、假设 2 和引理 2，有

$$|r_k-1|=\left|\frac{\mathrm{Ared}_k}{\mathrm{Pred}_k}-1\right|=\frac{\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|O(\|\boldsymbol{d}_k\|^2)+O(\|\boldsymbol{d}_k\|^4)}{\mathrm{Pred}_k}\leqslant$$

$$\frac{\|\boldsymbol{F}_k\|O(\|\boldsymbol{d}_k\|^2)+O(\|\boldsymbol{d}_k\|^4)}{O(\|\boldsymbol{F}_k\|\,\|\boldsymbol{d}_k\|)}=O(\|\boldsymbol{d}_k\|)\to 0.$$

再结合式(5)～(7)可得

$$\bar{r}_k=\frac{\overline{\mathrm{Ared}_k}}{\mathrm{Pred}_k}=\frac{\Lambda_k-\|\boldsymbol{F}_{k+1}\|^2}{\mathrm{Pred}_k}\geqslant\frac{\|\boldsymbol{F}_k\|^2-\|\boldsymbol{F}_{k+1}\|^2}{\mathrm{Pred}_k}=r_k\to 1.$$

所以 $r_k\to 1$，从而存在一个常数 $M>m$，使得对所有足够大的 $k$，都有 $\mu_k\leqslant M$.

根据文献[16]中定理 1 的证明结果，在不失一般性的情况下，假设对所有 $\bar{\boldsymbol{x}}\in N(\boldsymbol{x}^*,b_1)\cap X^*$ 都有秩 $\mathrm{rank}(\boldsymbol{J}(\boldsymbol{x}^*))=r$，假设 $\boldsymbol{J}(\bar{\boldsymbol{x}})$的奇异值分解为

$$\boldsymbol{J}(\bar{\boldsymbol{x}})=(\bar{\boldsymbol{U}}_1,\bar{\boldsymbol{U}}_2)\begin{pmatrix}\bar{\boldsymbol{\Sigma}}_1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\bar{\boldsymbol{V}}_1^{\mathrm{T}}\\ \bar{\boldsymbol{V}}_2^{\mathrm{T}}\end{pmatrix}=\bar{\boldsymbol{U}}_1\bar{\boldsymbol{\Sigma}}_1\bar{\boldsymbol{V}}_1^{\mathrm{T}},\tag{35}$$

其中 $\bar{\boldsymbol{\Sigma}}_1=\operatorname{diag}(\bar{\sigma}_1,\bar{\sigma}_2,\cdots,\bar{\sigma}_r)$，$\bar{\sigma}_1\geqslant\bar{\sigma}_2\geqslant\cdots\geqslant\bar{\sigma}_r>0$，并且 $\boldsymbol{U}=(\bar{\boldsymbol{U}}_1,\bar{\boldsymbol{U}}_2)$，$\boldsymbol{V}=(\bar{\boldsymbol{V}}_1,\bar{\boldsymbol{V}}_2)$是正交矩阵. 相应地，考虑 $\boldsymbol{J}(\bar{\boldsymbol{x}}_k)$的奇异值分解为

$$\boldsymbol{J}(\bar{\boldsymbol{x}}_k)=(\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3)\begin{pmatrix}\boldsymbol{\Sigma}_1 & 0 & 0\\ 0 & \boldsymbol{\Sigma}_2 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\boldsymbol{V}_1^{\mathrm{T}}\\ \boldsymbol{V}_2^{\mathrm{T}}\\ \boldsymbol{V}_3^{\mathrm{T}}\end{pmatrix}=\boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^{\mathrm{T}}+\boldsymbol{U}_2\boldsymbol{\Sigma}_2\boldsymbol{V}_2^{\mathrm{T}},\tag{36}$$

其中 $\boldsymbol{U}=(\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3)$，$\boldsymbol{V}=(\boldsymbol{V}_1,\boldsymbol{V}_2,\boldsymbol{V}_3)$是正交矩阵，$\boldsymbol{\Sigma}_1=\operatorname{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$，$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_r>0$，并且 $\boldsymbol{\Sigma}_2=\operatorname{diag}(\sigma_{r+1},\sigma_{r+2},\cdots,\sigma_{r+q})$，$\sigma_{r+1}\geqslant\sigma_{r+2}\geqslant\cdots\geqslant\sigma_{r+q}>0$.

**引理 4** 在假设 2 的条件下，对所有足够大的 $k$，有：

1) $\|\boldsymbol{U}_1\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{F}_k\|\leqslant O(\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|)$；

2) $\|\boldsymbol{U}_2\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{F}_k\|\leqslant O(\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2)$；

3) $\|\boldsymbol{U}_3\boldsymbol{U}_3^{\mathrm{T}}\boldsymbol{F}_k\|\leqslant O(\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2)$；

4) $\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|\leqslant O(\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2)$.

证明：结论 1)～3)的证明可参照文献[17]中引理 4.1 的证明得到，故略. 为证明结论 4)，由式(14)和矩阵摄动理论[18]，得

$$\|\operatorname{diag}(\boldsymbol{\Sigma}_1-\bar{\boldsymbol{\Sigma}}_1,\boldsymbol{\Sigma}_2,0)\|\leqslant\|\boldsymbol{J}_k-\boldsymbol{J}(\bar{\boldsymbol{x}}_k)\|\leqslant L_1\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|,$$

即

$$\|\boldsymbol{\Sigma}_1-\bar{\boldsymbol{\Sigma}}_1\|\leqslant L_1\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|,\qquad\|\boldsymbol{\Sigma}_2\|\leqslant L_1\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|.\tag{37}$$

因为$\{\boldsymbol{x}_k\}$收敛于解集 $X^*$，假设 $L_1\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|\leqslant\dfrac{\bar{\sigma}_r}{2}$对所有足够大的 $k$ 都成立，则由式(37)，有

$$\|\boldsymbol{\Sigma}_1^{-1}\|\leqslant\frac{1}{\bar{\sigma}_r-L_1\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|}\leqslant\frac{2}{\bar{\sigma}_r}.\tag{38}$$

再结合式(29)、引理 3 以及 $\boldsymbol{F}(\bar{\boldsymbol{x}}_k)=0$，可知 LM 参数满足

$$\lambda_k=\frac{\mu_k\|\boldsymbol{F}_k\|^2}{1+\|\boldsymbol{F}_k\|^2}\leqslant\mu_k\|\boldsymbol{F}_k\|^2\leqslant ML_2^2\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2.\tag{39}$$

由式(12)和式(36)，可得

$$\boldsymbol{d}_k=-\boldsymbol{V}_1(\boldsymbol{\Sigma}_1^2+\lambda_k\boldsymbol{I})^{-1}\boldsymbol{\Sigma}_1\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{F}_k-\boldsymbol{V}_2(\boldsymbol{\Sigma}_2^2+\lambda_k\boldsymbol{I})^{-1}\boldsymbol{\Sigma}_2\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{F}_k,$$

$$\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k=\boldsymbol{F}_k-\boldsymbol{U}_1\boldsymbol{\Sigma}_1(\boldsymbol{\Sigma}_1^2+\lambda_k\boldsymbol{I})^{-1}\boldsymbol{\Sigma}_1\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{F}_k-\boldsymbol{U}_2\boldsymbol{\Sigma}_2(\boldsymbol{\Sigma}_2^2+\lambda_k\boldsymbol{I})^{-1}\boldsymbol{\Sigma}_2\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{F}_k=$$

$$\lambda_k\boldsymbol{U}_1(\boldsymbol{\Sigma}_1^2+\lambda_k\boldsymbol{I})^{-1}\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{F}_k+\lambda_k\boldsymbol{U}_2(\boldsymbol{\Sigma}_2^2+\lambda_k\boldsymbol{I})^{-1}\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{F}_k+\boldsymbol{U}_3\boldsymbol{U}_3^{\mathrm{T}}\boldsymbol{F}_k.$$

再结合式(38)和式(39)，有

$$\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|\leqslant\lambda_k\|\boldsymbol{\Sigma}_1^{-2}\|\,\|\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{F}_k\|+\|\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{F}_k\|+\|\boldsymbol{U}_3\boldsymbol{U}_3^{\mathrm{T}}\boldsymbol{F}_k\|\leqslant$$

$$\frac{4L_2^2M\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^3}{\bar{\sigma}_r^2}+O(\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2)+O(\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2)=O(\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2).$$

**定理 2** 假设序列$\{\boldsymbol{x}_k\}$由 NALM 算法生成，则在假设 2 的条件下，序列$\{\boldsymbol{x}_k\}$二次收敛到问题(1)的解.

证明：由假设 2、引理 2 和引理 4 中 4)，有

$$c_1\|\bar{\boldsymbol{x}}_{k+1}-\boldsymbol{x}_{k+1}\|\leqslant\|\boldsymbol{F}(\boldsymbol{x}_{k+1})\|=\|\boldsymbol{F}(\boldsymbol{x}_k+\boldsymbol{d}_k)\|\leqslant$$

$$\|\boldsymbol{F}_k+\boldsymbol{J}_k\boldsymbol{d}_k\|+O(\|\boldsymbol{d}_k\|^2)=O(\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|^2).\tag{40}$$

另一方面，有

$$\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|=\operatorname{dist}(\boldsymbol{x}_k,X^*)\leqslant\|\bar{\boldsymbol{x}}_{k+1}-\boldsymbol{x}_k\|\leqslant\|\bar{\boldsymbol{x}}_{k+1}-\boldsymbol{x}_{k+1}\|+\|\boldsymbol{d}_k\|.$$

由引理 2，对足够大的 $k$ 有 $\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|\leqslant2\|\boldsymbol{d}_k\|\leqslant O(\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|)$. 因此，$\|\boldsymbol{d}_k\|=O(\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_k\|)$. 结合式(40)，可推出 $\|\boldsymbol{d}_{k+1}\|\leqslant O(\|\boldsymbol{d}_k\|^2)$，表明$\{\boldsymbol{x}_k\}$二次收敛到集合 $X^*$ 的一个解.

## 3 数值实验

下面将本文提出的 NALM 算法与非单调 LM(nonmonotone Levenberg-Marquardt，NLM)

算法[8]、自适应 LM(adaptive Levenberg-Marquardt，ALM)算法[19]的数值性能进行比较，以验证本文算法的有效性. 数值实验的测试函数集是通过修改文献[20]中给出的非奇异问题得到的，其形式与 Schnabel 等[21]构造的奇异测试问题形式相同.

所有算法程序均由 MATLAB 软件实现. 通过选择不同的 LM 参数考虑两种 ALM 算法：当选择 LM 参数中的 $\delta=1$ 时将算法记为 ALM1；当选择 $\delta=2$ 时将算法记为 ALM2. 在所有算法中，使用的参数为：$p_0=10^{-4}$，$p_1=0.25$，$p_2=0.75$，$N_0=5$，$\mu_0=0.01$，$m=10^{-8}$，$\sigma_1=0.25$，$\sigma_2=0.5$，$\beta_1=1.01$，$\beta_2=0.5$，$\beta_3=2$. 当满足 $\| \boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{F}_k \| \leqslant 10^{-6}$ 或者迭代次数超过 $100(n+1)$ 时算法停止迭代. 分别从 $-10\boldsymbol{x}_0$，$-\boldsymbol{x}_0$，$\boldsymbol{x}_0$，$10\boldsymbol{x}_0$ 和 $100\boldsymbol{x}_0$ 五个起始点运行测试问题，其中 $\boldsymbol{x}_0$ 由文献[21]给出.

本文利用 Dolan 等[22]提出的数值性能比较方法给出不同算法数值性能比较结果. 以迭代次数为例，在性能图中横坐标 $\tau$ 表示最佳比值因子，纵坐标 $P(\tau)$ 表示测试算法 $s$ 在最少迭代次数集的 $\tau$ 倍内所能解的问题个数占总问题个数的比率，它反映了测试算法迭代次数的数值性能. 图中最高曲线表示该测试算法在最少迭代次数集的 $\tau$ 倍内能求解的问题最多，因此性能图中位于右上角的算法数值性能更好. 不同算法的数值性能比较结果如图 1～图 4 所示.

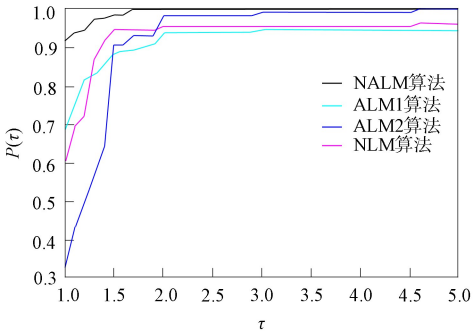



**图 1　不同算法迭代次数的性能比较**

**Fig. 1　Performance comparison for iterations of different algorithms**

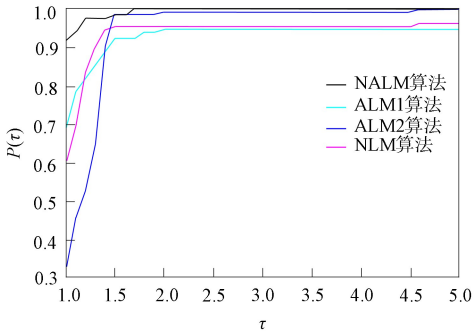



**图 2　不同算法函数计算次数的性能比较**

**Fig. 2　Performance comparison for function calculation times of different algorithms**

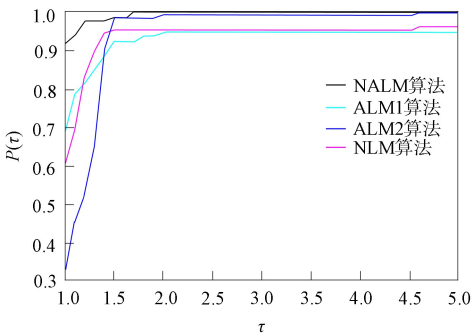



**图 3　不同算法梯度计算次数的性能比较**

**Fig. 3　Performance comparison for gradient calculation times of different algorithms**

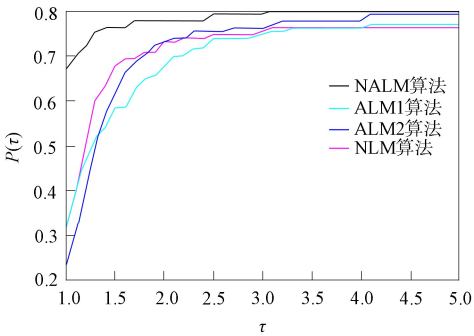



**图 4　不同算法 CPU 性能比较**

**Fig. 4　Performance comparison for CPU of different algorithms**

由图 1～图 3 可见：当 $\tau=5$ 时，NALM 算法和 ALM2 算法可成功求解所选的所有测试问题，而 NLM 算法只能成功求解约 96%的问题，ALM1 算法只能成功求解约 95%的问题；NALM 算法迭代次数、函数计算次数和梯度计算次数最少的问题约占总问题的 92%，而 ALM2 算法迭代次数、函数计算次数和梯度计算次数最少的问题约占总问题的 32%. 表明 NALM 算法比 ALM2 算法更有优势. 由图 4 可见，在解决本文所给的测试问题时，NALM 算法所用的运行时间相比其他算法更少. 实验结果表明，本文算法具有良好的数值性能，在求解非线性方程组问题方面性能更优.

综上所述，本文提出了一种新的非单调自适应加速 LM 算法，该算法基于一个有效的 LM 参数，利用非单调技术避免了迭代在局部跳跃，通过构造自适应函数提高了模型与目标函数的一致性. 数值

实验结果表明，本文所提算法不仅能快速收敛到解集，还能减少 Jacobi 矩阵计算次数，具有良好的发展前景.

## 参 考 文 献


[1] CHEN L, MA Y F. A New Modified Levenberg-Marquardt Method for Systems of Nonlinear Equations [J/OL]. J Math, (2023-02-10)[2023-06-16]. https://doi.org/10.1155/2023/6043780.

[2] SETYAWAN A, AQSA M K, SUSENO J E, et al. Gravity Inversion Modelling Using Simulated Annealing and the Levenberg-Marquardt Algorithm [J]. Inter J Geomate, 2022, 23(98): 109-116.

[3] HE X R, TANG J Y. A Smooth Levenberg-Marquardt Method without Nonsingularity Condition for wLCP [J]. AIMS Math, 2022, 7(5): 8914-8932.

[4] ZHAO J Y, ZHANG X J, ZHAO J L. A Levenberg-Marquardt Method for Tensor Approximation [J]. Symmetry, 2023, 15(3): 694-707.

[5] CHAMBERLAIN R M, POWELL M J D, LEMARECHAL C, et al. The Watchdog Technique for Forcing Convergence in Algorithm for Constrained Optimization [J]. Math Program Stud, 1982, 16(4): 1-17.

[6] GRIPPO L, LAMPARIELLO F, LUCIDI S. A Nonmonotone Line Search Technique for Newton's Method [J]. SIAM J Numer Anal, 1986, 23(4): 707-716.

[7] GRIPPO L, LAMPARIELLO F, LUCIDI S. A Truncated Newton Method with Nonmonotone Line Search for Unconstrained Optimization [J]. J Optim Theory Appl, 1989, 60(3): 401-419.

[8] AMINI K, ROSTAMI F, CARISTI G. An Efficient Levenberg-Marquardt Method with a New LM Parameter for Systems of Nonlinear Equations [J]. Optimization, 2018, 67(5): 637-650.

[9] KIMIAEI M. Nonmonotone Self-adaptive Levenberg-Marquardt Approach for Solving Systems of Nonlinear Equations [J]. Numer Func Anal Opt, 2017, 39(1): 47-66.

[10] HUANG B H, MA C F. A Shamanskii-Like Self-adaptive Levenberg-Marquardt Method for Nonlinear Equations [J]. Comput Math Appl, 2019, 77(2): 357-373.

[11] AMINI K, ROSTAMI F. A Modified Two Steps Levenberg-Marquardt Method for Nonlinear Equations [J]. J Comput Appl Math, 2015, 288: 341-350.

[12] HEI L. A Self-adaptive Trust Region Algorithm [J]. J Comput Math, 2003, 21(1): 229-236.

[13] WALMAG J M B, DELHEZ E J M. A Note on Trust-Region Radius Update [J]. SIAM J Optim, 2005, 16(2): 548-562.

[14] LU Y L, LI W Y, CAO M Y, et al. A Novel Self-adaptive Trust Region Algorithm for Unconstrained Optimization [J/OL]. J Appl Math, (2014-04-15)[2023-06-20]. https://doi.org/10.1155/2014/610612.

[15] POWELL M J D. Convergence Properties of a Class of Minimization Algorithms [J]. Nonlinear Program, 1975, 2(1): 1-27.

[16] BEHLING R, IUSEM A. The Effect of Calmness on the Solution Set of Systems of Nonlinear Equations [J]. Math Program, 2013, 137(1/2): 155-165.

[17] FAN J Y. A Modified Levenberg-Marquardt Algorithm for Singular System of Nonlinear Equations [J]. J Comput Math, 2003, 21(5): 625-636.

[18] STEWART G W, SUN J G. Matrix Perturbation Theory [J]. Appl Math Sci, 1991, 88(1): 1-15.

[19] ZHENG L, CHEN L, MA Y F. A Variant of the Levenberg-Marquardt Method with Adaptive Parameters for Systems of Nonlinear Equations [J]. AIMS Math, 2022, 7(1): 1241-1256.

[20] MORÉ J J, GARBOW B S, HILLSTROM K E. Testing Unconstrained Optimization Software [J]. ACM T Math Software, 1981, 7(1): 17-41.

[21] SCHNABEL R B, FRANK P D. Tensor Methods for Nonlinear Equations [J]. SIAM J Numer Anal, 1984, 21(5): 815-843.

[22] DOLAN E D, MORÉ J J. Benchmarking Optimization Software with Performance Profiles [J]. Math Program, 2002, 91(2): 201-213.


(责任编辑：赵立芹)