第55卷第5期
2015年9月

大连理工大学学报
Journal of Dalian University of Technology

Vol. 55, No. 5
Sept. 2015

应用数学

文章编号：1000-8608(2015)05-0548-05

求解半光滑方程组的LM方法收敛性分析

齐丽岩[1,2]， 肖现涛[1]， 张立卫[*1]

（1.大连理工大学 数学科学学院，辽宁 大连 116024；
2.大连海洋大学 理学院，辽宁 大连 116023）

摘要： Levenberg-Marquardt(LM)方法是一个经典并且有效的求解非线性方程组的方法，但是目前的研究都是针对光滑方程组的.在这样的背景下，研究求解半光滑非线性方程组的LM方法.构造了求解半光滑方程组的一个参数调整LM方法(S-PALM)，其中LM参数在每次迭代中是基于实际下降量和预测下降量的比值自动更新的.在水平有界的前提下，得到了S-PALM方法的全局收敛性.在强BD正则性成立的条件下，得到S-PALM方法的局部超线性收敛速度.

关键词： 半光滑方程组；Levenberg-Marquardt方法；全局收敛性；强BD正则性

中图分类号： O221.1 **文献标识码：** A **doi：** 10.7511/dllgxb201505016

0 引 言

考虑如下的非线性方程组：

$$\boldsymbol{F}(\boldsymbol{x})=\boldsymbol{0}$$

其中 $\boldsymbol{F}:\boldsymbol{R}^n\to\boldsymbol{R}^n$ 是连续可微的函数(光滑函数).定义效益函数

$$f(\boldsymbol{x}):=\frac{1}{2}\|\boldsymbol{F}(\boldsymbol{x})\|^2$$

那么

$$\nabla f(\boldsymbol{x})=\boldsymbol{J}\boldsymbol{F}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{F}(\boldsymbol{x})$$

由于 $f(\boldsymbol{x})$ 的最小值显然是0，求解非线性方程组等价于求解如下的无约束优化问题：

$$\min_{\boldsymbol{x}\in\boldsymbol{R}^n} f(\boldsymbol{x})$$

对迭代点 $\boldsymbol{x}_k$，定义

$$\nabla f_k=\nabla f(\boldsymbol{x}_k),\quad \boldsymbol{B}_k=\boldsymbol{J}\boldsymbol{F}^{\mathrm{T}}(\boldsymbol{x}_k)\boldsymbol{J}\boldsymbol{F}(\boldsymbol{x}_k)$$

求解非线性方程组的一个经典的方法是Levenberg-Marquardt(LM)方法[1-2].在第 k 次迭代，LM方法通过如下的公式得到下降方向 $\boldsymbol{p}_k$：

$$\boldsymbol{p}_k=-(\boldsymbol{B}_k+\mu_k\boldsymbol{I})^{-1}\nabla f_k$$

其中 $\mu_k>0$，是每次迭代中不断更新的参数.Levenberg-Marquardt方法是求解非线性方程组的一个非常流行并且有效的方法，但是大部分研究LM方法的相关文献都是考虑求解光滑的非线性方程组[3-5]，求解半光滑方程组的LM方法几乎没有研究.而半光滑方程组在优化领域中是非常重要的，许多优化问题和优化方法都与半光滑方程组的求解有关.因此，研究求解半光滑方程组的LM方法具有一定的理论意义和应用价值.

用 $\partial\boldsymbol{F}(\boldsymbol{x})$ 表示 $\boldsymbol{F}$ 在 $\boldsymbol{x}$ 处的广义雅克比(Clarke[6]意义下的广义雅克比).称函数 $\boldsymbol{F}$ 在 $\boldsymbol{x}$ 处是半光滑的(定义见文献[7])，如果 $\boldsymbol{F}$ 在 $\boldsymbol{x}$ 处是局部Lipschitz的并且对任意的 $\boldsymbol{h}\in\boldsymbol{R}^n$，$\lim\limits_{\boldsymbol{V}\in\partial\boldsymbol{F}(\boldsymbol{x}+t\boldsymbol{h}'),\boldsymbol{h}'\to\boldsymbol{h},t\downarrow 0}\{\boldsymbol{V}\boldsymbol{h}'\}$ 是存在的.在本文中，假设 $\boldsymbol{F}$ 是一个半光滑函数.这时 $\boldsymbol{p}_k$ 的公式无法使用.令 $\partial\boldsymbol{F}_k$ 为 $\boldsymbol{F}$ 在 $\boldsymbol{x}_k$ 处的广义雅克比，重新定义

$$f_k=f(\boldsymbol{x}_k),\quad \boldsymbol{F}_k=\boldsymbol{F}(\boldsymbol{x}_k),\quad \boldsymbol{g}_k=\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{F}_k,$$

$$\boldsymbol{B}_k=\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{V}_k\quad 其中\ \boldsymbol{V}_k\in\partial\boldsymbol{F}_k$$

那么也许可以用如下的公式替代前面 $\boldsymbol{p}_k$ 的公式：

$$\boldsymbol{p}_k=-(\boldsymbol{B}_k+\mu_k\boldsymbol{I})^{-1}\boldsymbol{g}_k$$

收稿日期： 2015-01-04； **修回日期：** 2015-07-28.
基金项目： 国家自然科学基金资助项目(11071029，11101064，91130007).
作者简介： 齐丽岩(1980-)，女，博士生，E-mail：qiliyan@dlou.edu.cn；张立卫*(1966-)，男，教授，博士生导师，E-mail：lwzhang@dlut.edu.cn.

从而得到求解半光滑非线性方程组的 LM 方法. 本文将考虑求解半光滑方程组的 LM 方法的局部和全局收敛性.

1　求解半光滑方程组的 LM 方法

考虑函数 f 在 $\boldsymbol{x}_k$ 附近的二次近似函数

$$m_k(\boldsymbol{p})=f_k+\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{p}+\frac{1}{2}\boldsymbol{p}^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{p}$$

对于方向 $\boldsymbol{p}_k$,定义

$$\rho_k=\frac{f_k-f(\boldsymbol{x}_k+\boldsymbol{p}_k)}{m_k(\boldsymbol{0})-m_k(\boldsymbol{p}_k)} \tag{1}$$

其中 $f_k-f(\boldsymbol{x}_k+\boldsymbol{p}_k)$称为第 k 次迭代的实际下降量,$m_k(\boldsymbol{0})-m_k(\boldsymbol{p}_k)$为预测下降量. 下面给出求解半光滑方程组的 LM 方法——半光滑方程组参数调整 LM 方法(parameter-adjusting LM method for semismooth equations,简称 S-PALM).

算法 1(S-PALM)　给定 $\boldsymbol{x}_0\in\boldsymbol{R}^n$,$\mu_0>0$:

for $k=0,1,2,\cdots$

　选取 $\boldsymbol{V}_k\in\partial\boldsymbol{F}(\boldsymbol{x}_k)$,计算

　　$\boldsymbol{p}_k=-(\boldsymbol{B}_k+\mu_k\boldsymbol{I})^{-1}\boldsymbol{g}_k$;

　利用式(1)计算 ρ_k;

　if　$\rho_k<1/4$

　　$\mu_{k+1}=4\mu_k$;

　else

　　if　$\rho_k>3/4$

　　$\mu_{k+1}=\mu_k/2$;

　　else

　　$\mu_{k+1}=\mu_k$;

　$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\boldsymbol{p}_k$;

end (for).

2　全局收敛性

先给出半光滑函数的一个性质.

引理 1[7]　若 $\boldsymbol{F}:\boldsymbol{R}^n\to\boldsymbol{R}^n$ 在 $\boldsymbol{x}$ 处是半光滑的,那么对任意的 $\boldsymbol{V}\in\partial\boldsymbol{F}(\boldsymbol{x})$有

$$\|\boldsymbol{F}(\boldsymbol{x}+\boldsymbol{p})-\boldsymbol{F}(\boldsymbol{x})-\boldsymbol{V}\boldsymbol{p}\|=O(\|\boldsymbol{p}\|^2)$$

引理 2[6]　若 $\boldsymbol{F}:\boldsymbol{R}^n\to\boldsymbol{R}^n$ 是 Lipschitz 连续的,那么 $\boldsymbol{F}$ 的次微分$\partial\boldsymbol{F}$ 这个集值映射是外半连续的(也叫上半连续).

下面的引理给出了预测下降量的一个下界,它在全局收敛性的证明中是一个关键.

引理 3

$$m_k(\boldsymbol{0})-m_k(\boldsymbol{p}_k)\geqslant\frac{1}{2}\|\boldsymbol{g}_k\|\min\left[\|\boldsymbol{p}_k\|,\frac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{B}_k\|}\right] \tag{2}$$

证明　容易观察到$(\boldsymbol{p},\lambda)=(\boldsymbol{p}_k,\mu_k)$满足下面的关系式:

$$\boldsymbol{g}_k+\boldsymbol{B}_k\boldsymbol{p}+\lambda\boldsymbol{p}=\boldsymbol{0},\ 0\leqslant\lambda\perp(\|\boldsymbol{p}_k\|-\|\boldsymbol{p}\|)\geqslant 0$$

而此关系式正好是下述优化问题的 KKT 系统:

$$\begin{aligned}&\min\quad m_k(\boldsymbol{p})\\&\text{s.t.}\quad \|\boldsymbol{p}\|^2\leqslant\|\boldsymbol{p}_k\|^2\end{aligned}$$

注意到 $m_k(\cdot)$是凸函数,于是得到 $\boldsymbol{p}_k$ 是上述优化问题的最优解.

采用文献[8]中定义的 Cauchy 步

$$\boldsymbol{p}_k^{\mathrm{C}}=\tau_k\|\boldsymbol{p}_k\|\frac{\boldsymbol{g}_k}{\|\boldsymbol{g}_k\|}$$

其中

$$\tau_k=\begin{cases}1; & \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{g}_k=\boldsymbol{0}\\ \min\left[1,\dfrac{\|\boldsymbol{g}_k\|^3}{\|\boldsymbol{p}_k\|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{g}_k}\right]; & \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{g}_k>\boldsymbol{0}\end{cases}$$

那么由文献[8]引理 4.3 可以得到下面的不等式:

$$m_k(\boldsymbol{0})-m_k(\boldsymbol{p}_k^{\mathrm{C}})\geqslant\frac{1}{2}\|\boldsymbol{g}_k\|\min\left[\|\boldsymbol{p}_k\|,\frac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{B}_k\|}\right]$$

注意到$\|\boldsymbol{p}_k^{\mathrm{C}}\|^2\leqslant\|\boldsymbol{p}_k\|^2$,那么

$$m_k(\boldsymbol{0})-m_k(\boldsymbol{p}_k)\geqslant m_k(\boldsymbol{0})-m_k(\boldsymbol{p}_k^{\mathrm{C}})$$

于是引理得证.

□

为了证明 S-PALM 的全局收敛性,需要下面的假设条件.

假设 1　设下面的水平集

$$L(\boldsymbol{x}_0)=\{\boldsymbol{x}\in\boldsymbol{R}^n:f(\boldsymbol{x})\leqslant f(\boldsymbol{x}_0)\}$$

是有界的,并且 $\boldsymbol{F}$ 在集合 $L(\boldsymbol{x}_0)$上是半光滑的.

下面给出算法 S-PALM 的全局收敛性,即证明算法生成的点列会收敛到效益函数 $f(\boldsymbol{x})$的一个稳定点.

定理 1　若假设 1 成立且$\{\boldsymbol{x}_k\}$由算法 1 生成,那么

$$\lim_{k\to\infty}\|\boldsymbol{g}_k\|=0 \tag{3}$$

证明　用反证法证明. 假设存在 $\varepsilon>0$ 和一个

正整数无穷序列 $\mathbf{N}$ 满足

$$\|\boldsymbol{g}_k\| \geqslant \varepsilon, \quad \forall k \in \mathbf{N} \tag{4}$$

对于 $\boldsymbol{x} \in L(\boldsymbol{x}_0)$，$\boldsymbol{p} \in \boldsymbol{R}^n$ 和 $\boldsymbol{V} \in \partial \boldsymbol{F}(\boldsymbol{x})$，定义

$$\boldsymbol{\Psi}(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{V}) = \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{V}\boldsymbol{p} - \boldsymbol{F}(\boldsymbol{x}+\boldsymbol{p})$$

$$\alpha(\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{V}) = \operatorname{dist}(\boldsymbol{V}, \partial \boldsymbol{F}(\boldsymbol{x}'))$$

与

$$\beta_1 = \max\{\|\boldsymbol{F}(\boldsymbol{x})\| : \boldsymbol{x} \in L(\boldsymbol{x}_0)\};$$

$$\beta_2 = \max\{\|\boldsymbol{V}\| : \boldsymbol{V} \in \partial \boldsymbol{F}(\boldsymbol{x}), \boldsymbol{x} \in L(\boldsymbol{x}_0)\};$$

$$\beta_3 = \max\{\|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\| : \boldsymbol{V} \in \partial \boldsymbol{F}(\boldsymbol{x}), \boldsymbol{x} \in L(\boldsymbol{x}_0)\}$$

由 ρ_k 的定义(1)可得

$$|\rho_k - 1| = \left| \frac{m_k(\boldsymbol{p}_k) - f(\boldsymbol{x}_k + \boldsymbol{p}_k)}{m_k(\boldsymbol{0}) - m_k(\boldsymbol{p}_k)} \right|$$

下面要估计 $|\rho_k - 1|$ 的一个界. 令 $\widehat{\boldsymbol{V}}_k$ 为 $\boldsymbol{V}_k$ 在集合 $\partial \boldsymbol{F}(\boldsymbol{x}_k + \boldsymbol{p}_k)$ 上的投影，记为

$$\widehat{\boldsymbol{V}}_k = \operatorname{Proj}_{\partial \boldsymbol{F}(\boldsymbol{x}_k + \boldsymbol{p}_k)}(\boldsymbol{V}_k)$$

由关系式

$$\boldsymbol{F}(\boldsymbol{x}_k + \boldsymbol{p}_k) = \boldsymbol{F}_k + \widehat{\boldsymbol{V}}_k \boldsymbol{p}_k + \boldsymbol{\Psi}(\boldsymbol{x}_k, \boldsymbol{p}_k, \widehat{\boldsymbol{V}}_k)$$

可得

$$\begin{aligned} f(\boldsymbol{x}_k + \boldsymbol{p}_k) = & \frac{1}{2}\|\boldsymbol{F}(\boldsymbol{x}_k + \boldsymbol{p}_k)\|^2 = \\ & f(\boldsymbol{x}_k) + \langle \boldsymbol{F}_k, \widehat{\boldsymbol{V}}_k \boldsymbol{p}_k + \boldsymbol{\Psi}(\boldsymbol{x}_k, \boldsymbol{p}_k, \widehat{\boldsymbol{V}}_k) \rangle + \\ & \frac{1}{2}\|\widehat{\boldsymbol{V}}_k \boldsymbol{p}_k + \boldsymbol{\Psi}(\boldsymbol{x}_k, \boldsymbol{p}_k, \widehat{\boldsymbol{V}}_k)\|^2 = \\ & m_k(\boldsymbol{p}_k) + \langle \boldsymbol{F}_k, (\widehat{\boldsymbol{V}}_k - \boldsymbol{V}_k)\boldsymbol{p}_k \rangle + \\ & \frac{1}{2}\boldsymbol{p}_k^{\mathrm{T}}[\widehat{\boldsymbol{V}}_k^{\mathrm{T}}\widehat{\boldsymbol{V}}_k - \boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{V}_k]\boldsymbol{p}_k + \\ & \langle \boldsymbol{F}_k, \boldsymbol{\Psi}(\boldsymbol{x}_k, \boldsymbol{p}_k, \widehat{\boldsymbol{V}}_k) \rangle + \\ & \langle \widehat{\boldsymbol{V}}_k \boldsymbol{p}_k, \boldsymbol{\Psi}(\boldsymbol{x}_k, \boldsymbol{p}_k, \widehat{\boldsymbol{V}}_k) \rangle + \\ & \frac{1}{2}\|\boldsymbol{\Psi}(\boldsymbol{x}_k, \boldsymbol{p}_k, \widehat{\boldsymbol{V}}_k)\|^2 \end{aligned}$$

由上面的等式可得

$$\begin{aligned} |m_k(\boldsymbol{p}_k) - f(\boldsymbol{x}_k + \boldsymbol{p}_k)| \leqslant & \beta_1 \alpha(\boldsymbol{x}_k, \boldsymbol{x}_k + \boldsymbol{p}_k, \boldsymbol{V}_k) \times \\ & \|\boldsymbol{p}_k\| + \beta_3 \|\boldsymbol{p}_k\|^2 + \\ & \beta_1 \|\boldsymbol{\Psi}(\boldsymbol{x}_k, \boldsymbol{p}_k, \widehat{\boldsymbol{V}}_k)\| + \\ & \beta_2 \|\boldsymbol{p}_k\| \times \\ & \|\boldsymbol{\Psi}(\boldsymbol{x}_k, \boldsymbol{p}_k, \widehat{\boldsymbol{V}}_k)\| + \\ & \frac{1}{2}\|\boldsymbol{\Psi}(\boldsymbol{x}_k, \boldsymbol{p}_k, \widehat{\boldsymbol{V}}_k)\|^2 \end{aligned} \tag{5}$$

通过引理3的不等式(2)可得，对所有的 $k \in \mathbf{N}$ 有

$$\begin{aligned} m_k(\boldsymbol{0}) - m_k(\boldsymbol{p}_k) \geqslant & \frac{1}{2}\|\boldsymbol{g}_k\| \times \\ & \min\left[\|\boldsymbol{p}_k\|, \frac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{B}_k\|}\right] \geqslant \\ & \frac{1}{2}\varepsilon \min\left[\|\boldsymbol{p}_k\|, \frac{\varepsilon}{\beta_3}\right] \end{aligned} \tag{6}$$

由于集值映射 $\partial \boldsymbol{F}$ 的外半连续性(引理2)，对任意的 $\boldsymbol{x} \in L(\boldsymbol{x}_0)$ 存在 $\delta_x > 0$ 使得

$$\partial \boldsymbol{F}(\boldsymbol{x}+\boldsymbol{p}) \subset \partial \boldsymbol{F}(\boldsymbol{x}) + \frac{\varepsilon}{40\beta_1}\boldsymbol{B}, \ \forall \|\boldsymbol{p}\| \leqslant \delta_x$$

其中 $\boldsymbol{B}$ 为 $\boldsymbol{R}^{n \times n}$ 上的单位闭球. 由于

$$L(\boldsymbol{x}_0) \subset \bigcup_{\boldsymbol{x} \in L(\boldsymbol{x}_0)} \mathbf{B}(\boldsymbol{x}, \delta_x)$$

其中

$$\mathbf{B}(\boldsymbol{x}, \delta_x) = \{\boldsymbol{x}' \in \boldsymbol{R}^n : \|\boldsymbol{x}' - \boldsymbol{x}\| < \delta_x\}$$

可得 $\{\mathbf{B}(\boldsymbol{x}, \delta_x) : \boldsymbol{x} \in L(\boldsymbol{x}_0)\}$ 是集合 $L(\boldsymbol{x}_0)$ 的一个开覆盖. 因为 $L(\boldsymbol{x}_0)$ 是非空紧致的，由有限开覆盖定理可得，存在集合 $L(\boldsymbol{x}_0)$ 中有限个点(不妨记为 $y_1, \cdots, y_q$)使得

$$L(\boldsymbol{x}_0) \subset \bigcup_{j=1}^{q} \mathbf{B}(y_j, \delta_{y_j})$$

令 $\delta_0 = \min\{\delta_{y_j} : j = 1, \cdots, q\}$，那么对 $\forall \|\boldsymbol{p}\| \leqslant \delta_0$，$\forall \boldsymbol{x} \in L(\boldsymbol{x}_0)$ 有

$$\partial \boldsymbol{F}(\boldsymbol{x}+\boldsymbol{p}) \subset \partial \boldsymbol{F}(\boldsymbol{x}) + \frac{\varepsilon}{40\beta_1}\boldsymbol{B} \tag{7}$$

注意到 $\boldsymbol{F}$ 在 $L(\boldsymbol{x}_0)$ 上是半光滑的，由引理1可得 $\|\boldsymbol{\Psi}(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{V})\| = \|\boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{V}\boldsymbol{p} - \boldsymbol{F}(\boldsymbol{x}+\boldsymbol{p})\| = O(\|\boldsymbol{p}\|^2)$

由 $L(\boldsymbol{x}_0)$ 的紧致性，通过同样的分析，结合 $\boldsymbol{F}$ 在 $L(\boldsymbol{x}_0)$ 上是半光滑的，存在 $\delta_1 > 0$ 使得对 $\forall \|\boldsymbol{p}\| \leqslant \delta_1$，$\forall \boldsymbol{x} \in L(\boldsymbol{x}_0)$ 有

$$\frac{\|\boldsymbol{\Psi}(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{V})\|}{\|\boldsymbol{p}\|} \leqslant \min\left[\frac{\varepsilon}{40\beta_1}, \sqrt{\frac{\varepsilon}{20\|\boldsymbol{p}\|}}\right] \tag{8}$$

和

$$\|\boldsymbol{\Psi}(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{V})\| \leqslant \frac{\varepsilon}{40\beta_2} \tag{9}$$

由 $\boldsymbol{g}_k = \boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{F}_k$ 可得 $\|\boldsymbol{g}_k\| \leqslant \beta_1\beta_2$. 由 $\lambda_{\min}(\boldsymbol{B}_k + \mu_k\boldsymbol{I}) \geqslant \mu_k$ 可得，$\|(\boldsymbol{B}_k + \mu_k\boldsymbol{I})^{-1}\|_2 = [\lambda_{\min}(\boldsymbol{B}_k + \mu_k\boldsymbol{I})]^{-1} \leqslant [\mu_k]^{-1}$，因此

$$\|\boldsymbol{p}_k\| = \|(\boldsymbol{B}_k + \mu_k\boldsymbol{I})^{-1}\boldsymbol{g}_k\| \leqslant [\mu_k]^{-1}\beta_1\beta_2$$

令 N 为 $\mathbf{N}$ 中第一个指标. 考虑充分大的 $\bar{\mu} > \mu_N$ 使得对任意的 $k \in \mathbf{N}$，$\mu_k \geqslant \bar{\mu}$，都有 $\boldsymbol{p}_k = -(\boldsymbol{B}_k +$

$\mu_k \boldsymbol{I})^{-1}\boldsymbol{g}_k$ 充分小并且满足

$$\|\boldsymbol{p}_k\| \leqslant \min\left[\frac{\varepsilon}{40\beta_3}, \delta_0, \delta_1\right]$$

那么由式(7)～(9)可得

$$\alpha(\boldsymbol{x}_k, \boldsymbol{x}_k+\boldsymbol{p}_k, \boldsymbol{V}_k) \leqslant \frac{\varepsilon}{40\beta_1};$$

$$\|\boldsymbol{p}_k\| \leqslant \frac{\varepsilon}{40\beta_3};$$

$$\|\boldsymbol{\Psi}(\boldsymbol{x}_k, \boldsymbol{p}_k, \boldsymbol{V}_k)\| \leqslant \min\left[\frac{\varepsilon}{40\beta_2}, \frac{\varepsilon\|\boldsymbol{p}_k\|}{40\beta_1}, \sqrt{\frac{\varepsilon\|\boldsymbol{p}_k\|}{20}}\right] \tag{10}$$

由式(5)和(10)可得

$$|m_k(\boldsymbol{p}_k)-f(\boldsymbol{x}_k+\boldsymbol{p}_k)| \leqslant \frac{1}{8}\varepsilon\|\boldsymbol{p}_k\|$$

由式(6)可得，对 $\boldsymbol{p}_k, k\in\mathbf{N}$ 充分小有

$$|\rho_k-1| = \frac{|m_k(\boldsymbol{p}_k)-f(\boldsymbol{x}_k+\boldsymbol{p}_k)|}{|m_k(\boldsymbol{0})-m_k(\boldsymbol{p}_k)|} \leqslant \frac{\frac{1}{8}\varepsilon\|\boldsymbol{p}_k\|}{\frac{1}{2}\varepsilon\min\left[\|\boldsymbol{p}_k\|, \frac{\varepsilon}{\beta_3}\right]} \leqslant \frac{1}{4}$$

因此，$\rho_k \geqslant 3/4$. 那么由算法 1 的结构，当 $k\in\mathbf{N}$，$\mu_k \geqslant \bar{\mu}$，可得 $\mu_{k+1} \leqslant \mu_k$.

这样得到 μ_k 增加的情况(变大 4 倍)只能在 $\mu_k < \bar{\mu}$ 时发生. 因此，得到

$$\mu_k \leqslant 4\bar{\mu}, \forall k\in\mathbf{N} \tag{11}$$

现在假设存在一个无穷序列 $\mathscr{K}\subset\mathbf{N}$ 使得对任意的 $k\in\mathscr{K}$ 都有 $\rho_k \geqslant 1/4$. 若 $k\in\mathscr{K}$，由式(2)可得

$$f(\boldsymbol{x}_k)-f(\boldsymbol{x}_{k+1}) \geqslant \frac{1}{4}[m_k(\boldsymbol{0})-m_k(\boldsymbol{p}_k)] \geqslant \frac{1}{8}\varepsilon\min[\|\boldsymbol{p}_k\|, \varepsilon/\beta_3]$$

由于假设 f 是下有界的，那么由上面的不等式可得

$$\lim_{k\in\mathscr{K}, k\to\infty}\|\boldsymbol{p}_k\| = 0$$

这意味着 $\mu_k \to \infty$，$k\in\mathscr{K}$，$k\to\infty$，但这与式(11)矛盾. 即那样的无穷序列 $\mathscr{K}$ 是不存在的，所以当 k 充分大时必有 $\rho_k < 1/4$. 在这种情况下，μ_k 最终每一次迭代都会增大 4 倍，于是 $\mu_k \to \infty$，这与式(11)矛盾. 因此，式(4)是不正确的，定理得证. □

3　局部收敛性

讨论在强 BD 正则条件下算法 S-PALM 的局部收敛速度. 令 $\boldsymbol{X}^*$ 表示半光滑方程组的解集. 首先给出如下的定义.

定义 1(强 BD 正则性)　对于 $\boldsymbol{x}^*\in\boldsymbol{X}^*$，称强 BD 正则性在 $\boldsymbol{x}^*$ 处成立，如果所有的 $\boldsymbol{V}\in\partial\boldsymbol{F}(\boldsymbol{x}^*)$ 是非奇异的.

引理 4　假设 $\{\boldsymbol{x}_k\}$ 是算法 1 生成的点列，则

$$\lim_{k\to\infty}\mu_k = 0$$

证明　利用定理 1 证明中的分析，可以得到，存在常数 $c>0$ 使得当 k 充分大时有

$$|\rho_k-1| \leqslant c\|\boldsymbol{p}_k\|$$

由算法的全局收敛性可知 $\lim\limits_{k\to\infty}\boldsymbol{p}_k = \boldsymbol{0}$. 于是，当 k 充分大时总有 $\rho_k > 3/4$. 所以，当 k 充分大时，总有 μ_k 是下降的(1/2 倍的速度)，所以 $\lim\limits_{k\to\infty}\mu_k = 0$.

□

引理 5　若强 BD 正则性在 $\boldsymbol{x}^*$ 处成立，那么存在 $\boldsymbol{x}^*$ 的邻域 $N(\boldsymbol{x}^*)$ 和常数 C，使得对所有的 $\boldsymbol{y}\in N(\boldsymbol{x}^*)$ 和任意的 $\boldsymbol{V}\in\partial\boldsymbol{F}(\boldsymbol{y})$，有 $\boldsymbol{V}$ 是非奇异的且

$$\|\boldsymbol{V}^{-1}\| \leqslant C$$

定理 2　对于 $\boldsymbol{x}^*\in\boldsymbol{X}^*$，假设强 BD 正则性在 $\boldsymbol{x}^*$ 处成立，那么算法 1 产生的点列 $\{\boldsymbol{x}_k\}$ 在 $\boldsymbol{x}^*$ 的邻域内超线性收敛到 $\boldsymbol{x}^*$.

证明　假设算法 1 在 $k=0$ 时 $\boldsymbol{x}_0$ 是在 $\boldsymbol{x}^*$ 的某个邻域内，那么由引理 4 和 5

$$\begin{aligned}
\mathrm{dist}(\boldsymbol{x}_{k+1}, \boldsymbol{X}^*) &= \|\boldsymbol{x}_{k+1}-\boldsymbol{x}^*\| = \|\boldsymbol{x}_k-\boldsymbol{x}^*+\boldsymbol{p}_k\| = \\
&\|\boldsymbol{x}_k-\boldsymbol{x}^*-(\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{V}_k+\mu_k\boldsymbol{I})^{-1}\times \\
&\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{F}_k\| = \\
&\|(\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{V}_k+\mu_k\boldsymbol{I})^{-1}((\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{V}_k+ \\
&\mu_k\boldsymbol{I})(\boldsymbol{x}_k-\boldsymbol{x}^*)-\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{F}_k)\| \leqslant \\
&\|(\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{V}_k+\mu_k\boldsymbol{I})^{-1}\|\|\boldsymbol{V}_k\|\times \\
&\|-\boldsymbol{F}_k+\boldsymbol{V}_k(\boldsymbol{x}_k-\boldsymbol{x}^*)\|+ \\
&\|(\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{V}_k+\mu_k\boldsymbol{I})^{-1}\|\times \\
&\|\mu_k(\boldsymbol{x}_k-\boldsymbol{x}^*)\| \leqslant \\
&\|(\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{V}_k)^{-1}\|\|\boldsymbol{V}_k\|\times \\
&\|\boldsymbol{F}(\boldsymbol{x}^*)-\boldsymbol{F}_k-\boldsymbol{V}_k(\boldsymbol{x}^*-\boldsymbol{x}_k)\|+ \\
&\mu_k\|(\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{V}_k)^{-1}\|\|\boldsymbol{x}_k-\boldsymbol{x}^*\| = \\
&o(\|\boldsymbol{x}_k-\boldsymbol{x}^*\|) = o(\mathrm{dist}(\boldsymbol{x}_k, \boldsymbol{X}^*))
\end{aligned}$$

□

4 结论与展望

本文考虑了求解半光滑非线性方程组的一类 Levenberg-Marquardt 方法. 在水平有界的条件下,证明了算法的全局收敛性. 在强 BD 正则的条件下,讨论了算法的局部超线性收敛速度. 但是,强 BD 正则的假设条件一般情况下不容易满足,因此应该讨论在一个更弱的假设条件下算法的局部收敛性. 对于求解光滑非线性方程组,文献[3-4]证明了在局部误差有界的条件下,LM 方法是局部二次收敛的. 则在局部误差有界的条件下,半光滑的 LM 方法是否也具有局部二次收敛速度是值得研究的课题.

参考文献:

[1] Levenberg K. A method for the solution of certain nonlinear problems in least squares [J]. **Quarterly Journal of Applied Mathematics**, 1944, **2**(2): 164-166.

[2] Marquardt D M. An algorithm for least-squares estimation of nonlinear inequalities [J]. **Journal of the Society for Industrial and Applied Mathematics**, 1963, **11**(2): 431-441.

[3] Yamashita N, Fukushima M. On the rate of convergence of the Levenberg-Marquardt method [J]. **Computing Supplementa**, 2001, **15**: 239-249.

[4] FAN Jin-yan, YUAN Ya-xiang. On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption [J]. **Computing**, 2005, **74**(1): 23-39.

[5] Fan J. The modified Levenberg-Marquardt method for nonlinear equations with cubic convergence [J]. **Mathematics of Computation**, 2012, **81**(277): 447-466.

[6] Clarke F H. **Optimization and Nonsmooth Analysis** [M]. New York: Wiley, 1983.

[7] Qi L, Sun J. A nonsmooth version of Newton's method [J]. **Mathematical Programming**, 1993, **58**(3): 353-367.

[8] Nocedal J, Wright S J. **Numerical Optimization** [M]. New York: Springer, 2006.

Convergence analysis of LM method for semismooth equations

QI Li-yan[1,2], XIAO Xian-tao[1], ZHANG Li-wei[*1]

(1. School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China;
2. School of Sciences, Dalian Ocean University, Dalian 116023, China)

Abstract: Levenberg-Marquardt (LM) method is a classical and very efficient method for solving nonlinear equations. However, most of the references on LM method considered the smooth equations. Based on this observation, it is interesting to study the LM method for semismooth equations. A parameter-adjusting LM method for semismooth equations (S-PALM) is constructed to solve semismooth nonlinear equations, in which the parameter is updated based on the ratio between actual reduction and predicted reduction. Under level bounded condition, the global convergence of S-PALM is proved. Under strong BD regularity assumption, the local superlinear convergence rate of S-PALM is established.

Key words: semismooth equations; Levenberg-Marquardt method; global convergence; strong BD regularity