研究简讯

求解一类非光滑优化问题的 Gauss-Newton 法 *

李 冲　王兴华
(东南大学应用数学系,南京 210096)　(杭州大学数学和信息科学系,杭州 310028)

摘要　对一般的凸函数建立了求解复合凸优化问题的 Gauss-Newton 法的局部二阶收敛性,从而在本质上推广了 Burke 等人的结果.

关键词　Gauss-Newton 法　弱锐最优解集　正则点　二阶收敛性

求解非线性方程组最小二乘解的 Gauss-Newton 法是上世纪初由 Gauss 提出的,现在已被广泛地用来求解下面的复合凸优化问题:

$$(\mathrm{P})\ \min f(x) := h(F(x)), \tag{1}$$

其中 $h:\mathbb{R}^m\to\mathbb{R}$ 是凸函数 $F:\mathbb{R}^n\to\mathbb{R}^m$ 是具有连续导数的函数. 复合凸优化问题近年来一直受到广泛关注,其原因是它包含诸如非线性包含、minimax 问题以及罚函数技巧等非光滑优化理论中许多重要的问题和技巧,同时也提供了一个新的统一框架,使优化问题数值解的理论分析得到别开生面的发展. 因此目前它已成为非线性优化理论中的一个主流方向.

1985 年,Womersley[1]在关于解的强唯一前提下证明了 Gauss-Newton 法的局部二阶收敛性定理,从而将 Jittorntrum 和 Osborne[2]在 h 为一范数时的结果推广到一般的凸函数. 然而,Womersley 的条件只能保证 Gauss-Newton 序列收敛到(P)的一个局部最优解. 最近 Burke 和 Ferris[3]在 Gauss-Newton 法的收敛性研究上取得了重要进展,其突出的特点是不必要求问题的解集是单点集,事实上甚至可以允许解集为无界. 他们的工作主要基于下列两个假设:

(ⅰ)h 的最优解集 $\mathbb{C}$ 是弱锐最优解集. 即存在 $\lambda>0$ 使得 $\forall y\in\mathbb{R}^m$ 满足 $h(y)\geqslant h_{\min}+\lambda d(y,\mathbb{C})$,其中 $h_{\min}=\min_y h(y)$,而 $d(y,\mathbb{C})$则表示点 $y\in\mathbb{R}^m$ 到集合 $\mathbb{C}$ 的距离.

(ⅱ)关于下列包含关系的正则点存在:

$$F(x)\in\mathbb{C}, \tag{2}$$

即存在 $\bar{x}\in\mathbb{R}^m$,使得

$$\ker(F'(\bar{x})^{\mathrm{T}})\cap\Gamma_{\mathbb{C}}(F(\bar{x}))=\{0\},$$

其中集值映射 $\Gamma_{\mathbb{C}}:\mathbb{R}^m\to\mathbb{R}^m$ 定义如下:

$$\Gamma_{\mathbb{C}}(y)=(\mathrm{cone}(\mathbb{C}-y))^0=\{x^*\in\mathbb{R}^m:\langle x^*,x\rangle\leqslant 1,\ \forall x\in\mathrm{cone}(\mathbb{C}-y)\},\ \forall y\in\mathbb{R}^m.$$

在此前提下,他们建立了 Gauss-Newton 序列向(P)的整体最优解的局部二阶收敛性. 本文

1999-08-27 收稿,1999-10-25 收修改稿

* 国家自然科学基金(批准号:19971013)和江苏省自然科学基金(批准号:BK99001)资助项目,同时也得到国家重点基础研究专项经费的部分资助

在 Burke 等人上述工作的基础上继续进行这方面的研究. 出乎预料的是,我们发现 Gauss-Newton 序列的收敛性竟基本上与凸函数 h 的附加性质无关. 事实上,本文对所有的凸函数 h 建立了 Gauss-Newton 方法的局部二阶收敛性. 进一步提出了松弛型 Gauss-Newton 法,并给出其超线性收敛性.

1 Gauss-Newton 法及其收敛性定理

对 $\Delta>0$ 和 $x\in\mathbb{R}^n$,令 $D_\Delta(x)$表示下述优化问题

$$\min\{h(F(x)+F'(x)d):\|d\|\leqslant\Delta\}\tag{3}$$

的解集.

算法 1 设 $\eta\geqslant1$, $\Delta\in(0,+\infty]$, $x^0\in\mathbb{R}^n$. 对于 $k=1,2,\cdots$, 若 x^k 已算出,则

(i)如果 $h(F(x^k))=\min\{h(F(x^k)+F'(x^k)d):\|d\|\leqslant\Delta\}$,算法终止;否则,取 $d^k\in D_\Delta(x^k)$使之满足 $\|d^k\|\leqslant\eta d(0,D_\Delta(x^k))$.

(ii)令 $x^{k+1}=x^k+d^k$.

令 B 为$\mathbb{R}^m$ 或$\mathbb{R}^n$ 的闭单位球,而 $x+rB$ 则表示以 x 为中心 r 为半径的闭球. Burke 和 Ferris 在文献[3]中证明了下面的定理.

定理 设$\mathbb{C}$ 是 h 的弱锐最优解集,$\bar{x}\in\mathbb{R}^n$ 是包含关系(2)的正则点,设 $0<\bar{\delta}<\Delta$,使得

$$d(0,D_\Delta(x))\leqslant\beta d(F(x),\mathbb{C})\text{ 和}\{d\in\mathbb{R}^n:\|d\|\leqslant\Delta,\ F(x)+F'(x)d\in\mathbb{C}\}\neq\Phi,\tag{4}$$

在 $\bar{x}+\bar{\delta}B$ 上对某个 β 成立(由文献[3]中的命题 3.3 知,这样的 $\bar{\delta}$ 是存在的). 假设 F' 在 $\bar{x}+\bar{\delta}B$ 上满足 Lipschitz 常数为 L 的 Lipschitz 条件,而 h 在 $F(\bar{x}+\bar{\delta}B)+1/8LB$ 上满足 Lipschtiz 常数为 M 的 Lipschitz 条件. 若存在 $\delta>0$,使得

a) $\delta<\min\{\bar{\delta}/2,1\}$;

b) $d(F(\bar{x}),C)<\delta/(2\eta\beta)$;

c) $\theta:=\eta LM\delta\beta/\lambda<1$.

则存在 $\bar{x}$ 的邻域$M(\bar{x})$,使 $\forall\,x^0\in M(\bar{x})$,以 x^0 为初值的 Gauss-Newton 序列$\{x^k\}$二阶收敛到(P)的整体最优解 x^*.

值得指出的是弱锐最优解集假设是非常强的. 事实上,对任意凸函数 f,若 $s>1$ 则 $h(y)=(f(y)-f_{\min})^s$ 不存在弱锐最优解集. 另一类不存在弱锐最优解集的函数类是 Gateaux 可微凸函数类. 因此,在 Gauss-Newton 法的收敛性定理中解除弱锐最优解集的存在性假设是非常有意义的. 我们有

定理 1 设$\mathbb{C}$ 是 h 的最优解集,$\bar{x}\in\mathbb{R}^n$ 是包含关系(2)式的正则点. 设 $0<\bar{\delta}<\Delta$,使得(4)式在 $\bar{x}+\bar{\delta}B$ 上对某个 β 成立. 假设 F' 在 $\bar{x}+\bar{\delta}B$ 上满足 Lipschitz 常数为 L 的 Lipschitz 条件. 若存在 $\delta>0$,使得

a) $\delta<\min\{\bar{\delta}/2,1\}$;

b) $d(F(\bar{x}),\mathbb{C})<\delta/(2\eta\beta))$;

c) $\eta L\delta\beta<2$.

则存在 $\bar{x}$ 的邻域$M(\bar{x})$,使对 $\forall\,x^0\in M(\bar{x})$,以 x^0 为初值的 Gauss-Newton 序列$\{x^k\}$二阶收敛到

(P)的整体最优解 x^*.

2 松弛型 Gauss-Newton 法及其收敛性定理

考虑到数值计算的实际背景,下面给出的松弛型 Gauss-Newton 法及其超线性收敛性具有重要的应用价值.

以 $D_{\Delta}^{k}(x^{k})$表示所有满足 $\| d \| \leqslant \Delta$ 和

$$h(F(x^k)+F'(x^k)d) \leqslant \min\{h(F(x^k)+F'(x^k)d):\| d \| \leqslant \Delta\}+\| d^{k-1}\|^{\alpha}$$

的 $d\in\mathbb{R}^n$ 全体.

算法2 设 $\eta\geqslant 1$, $\alpha>1$, $\Delta\in(0,+\infty]$, $x^0\in\mathbb{R}^n$ 及 $d^{-1}\in\mathbb{R}^n$, $d^{-1}\neq 0$. 对于 $k=1,2,\cdots$, 若 x^k 已算出,则

ⅰ) 如果 $h(F(x^k))=\min\{h(F(x^k)+F'(x^k)d):\| d \| \leqslant \Delta\}+\| d^{k-1}\|^{\alpha}$,令 $d^k=\| d^{k-1}\|^{\alpha}d^{k-1}$;否则,取 $d^k\in D_{\Delta}^{k}(x^k)$,使之满足 $\| d^k \| \leqslant \eta d(0,D_{\Delta}^{k}(x^k))$.

ⅱ) 令 $x^{k+1}=x^k+d^k$.

定理2 设$\mathbb{C}$是 h 的弱锐最优解集,$\bar{x}\in\mathbb{R}^n$ 是包含关系(2)式的正则点. 设 $0<\bar{\delta}<\Delta$,使得(4)式在 $\bar{x}+\bar{\delta}B$ 上对某个 β 成立. 假设 F' 在 $\bar{x}+\bar{\delta}B$ 上满足 Lipschitz 常数为 L 的 Lipschitz 条件. 若存在 $\delta>0$,使得

a) $\delta<\min\{\bar{\delta}/(1+c),1\}$;

b) $d(F(\bar{x}),\mathbb{C})<\delta/(2\eta\beta)$;

c) $\eta L\delta\beta/2+\eta\beta\delta^{\alpha-1}/\lambda<1$.

其中 $p=\min\{2,\sqrt{\alpha}\}$, $c=\sum_{i=0}^{+\infty}(1/2)^{p^i}$. 则存在 $\bar{x}$ 的邻域$M(\bar{x})$,使$\forall x^0\in M(\bar{x})$,以 x^0 为初值的 Gauss-Newton 序列$\{x^k\}p$ 阶收敛到(P)的整体最优解 x^*.

注 比较定理1和2,则发现定理2需要假设$\mathbb{C}$是 h 的弱锐最优解集. 注意到算法2中的校正项 d^k 是优化问题(3)式的一个逼近解,故定理2中的弱锐最优解集要求是自然的.

参 考 文 献

1 Womersley R S. Local properties of algorithms for minimizing nonsmooth composite functions. Mathematical Programming, 1985, 32: 69

2 Jittorntrum K, Osborne M R. Strong uniqueness and second order convergence in nonlinear discrete approximation. Numerische Mathematik, 1980, 34: 439

3 Burke J V, Ferris M C. A Gauss-Newton method for convex composite optimization. Mathematical Programming, 1995, 71: 179