一种无记忆拟牛顿法的收敛性

颜世建

(南京师范大学数学与计算机科学学院,210097,江苏,南京)

[摘要] 在 $f(x)$ 为二阶连续可微凸函数的条件下,证明了一种无记忆拟牛顿法的收敛性.

[关键词] 凸函数,无记忆拟牛顿法,收敛性

[中图分类号]O241, [文献标识码]A, [文章编号]1001-4616(2004)02-0016-03

0 引言

对无约束最优化问题,文[1]提出一种无记忆拟牛顿法,算法步骤如下.

算法 1

1° 初始点 $x^{(0)}, d^{(0)} = -g^{(0)} = -\nabla f(x^{(0)}), \varepsilon > 0, k = 0$.

2° 若 $\| g^{(k)} \|_2 \leqslant \varepsilon, x^* \approx x^{(k)}$,停算;否则用某种线搜索准则求 $\alpha_k > 0$,置 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$.

3° 计算 $s_k = x^{(k+1)} - x^{(k)}, g^{(k+1)} = \nabla f(x^{(k+1)}), y_k = g^{(k+1)} - g^{(k)}$.

若 $s_k^T y_k = \| s_k \|_2 \| y_k \|_2$,置 $d^{(k+1)} = -\frac{\| s_k \|_2}{\| y_k \|_2} g^{(k+1)}, k = k+1$,转 2°;否则转 4°.

4° 计算

$$l^{(k)} = \frac{\| y_k \|_2 s_k + \| s_k \|_2 y_k}{\| \| y_k \|_2 s_k + \| s_k \|_2 y_k \|_2}, \qquad u^{(k)} = \frac{\| s_k \|_2 l^{(k)} - s_k}{\| \| s_k \|_2 l^{(k)} - s_k \|_2}, \qquad v^{(k)} = \frac{\| y_k \|_2 l^{(k)} - y_k}{\| \| y_k \|_2 l^{(k)} - y_k \|_2},$$

$$w^{(k)} = g^{(k+1)} - 2(v^{(k)^T} g^{(k+1)}) v^{(k)}, \qquad \hat{s}^{(k)} = w^{(k)} - 2(u^{(k)^T} w^{(k)}) u^{(k)}, \qquad d^{(k+1)} = -\frac{\| s_k \|_2}{\| y_k \|_2} \hat{s}^{(k)},$$

置 $k = k+1$,转 2°.

文[1]还在 $f(x)$ 为二阶连续可微一致凸函数且采用精确线搜索或渐近精确线搜索的条件下证明了算法 1 的收敛性.本文改进了文[1]的收敛性定理,在 $f(x)$ 为二阶连续可微凸函数及采用 wolfe 线搜索准则的条件下,证明了算法 1 的收敛性.

1 收敛性定理

记 $H_1^{(k)} = I - 2u^{(k)} u^{(k)^T}$, $H_2^{(k)} = I - 2v^{(k)} v^{(k)^T}$, 当 $s_k^T y_k = \| s_k \|_2 \| y_k \|_2$ 时,令 $H_{k+1} = \frac{\| s_k \|_2}{\| y_k \|_2} I$,由于 $\| y_k \|_2 s_k = \| s_k \|_2 y_k$,有 $H_{k+1} y_k = s_k$,显然 $d^{(k+1)} = -H_{k+1} g^{(k+1)}$.当 $s_k^T y_k < \| s_k \|_2 \| y_k \|_2$ 时,令 $H_{k+1} = \frac{\| s_k \|_2}{\| y_k \|_2} H_1^{(k)} H_2^{(k)}$,由于 $H_1^{(k)} H_2^{(k)} y_k = \| y_k \|_2 H_1^{(k)} l^{(k)} = \frac{\| y_k \|_2}{\| s_k \|_2} s_k$,故有 $H_{k+1} y_k = s_k$,另 $d^{(k+1)} = -\frac{\| s_k \|_2}{\| y_k \|_2} \hat{s}^{(k)} = -H_{k+1} g^{(k+1)}$.因而算法 1 是一种无记忆拟牛顿法.

定理 1 设 $f(x)$ 为 $\mathbf{R}^n$ 上二阶连续可微的凸函数,水平集 $L(x^{(0)}) = \{ x \mid f(x) \leqslant f(x^{(0)}), x \in \mathbf{R}^n \}$ 有界.算法 1 中采用 wolfe 线搜索准则,则存在 $M \geqslant 1$ 使

$$-g^{(k)^T} d^{(k)} \geqslant \frac{1}{M} \| g^{(k)} \|_2^2, \forall k \geqslant 0.$$

证明 由于 $\{ f(x^{(k)}) \}$ 单调递减,故 $\{ x^{(k)} \} \subset L(x^{(0)})$,由 $L(x^{(0)})$ 有界,$f(x)$ 为二阶连续可微的凸函数

收稿日期: 2003-10-10.

作者简介: 颜世建,1946-,南京师范大学数学与计算机科学学院教授,主要从事计算数学的教学与研究,E-mail:yanshijian@163.com

可知存在 $M \geqslant 1$ 使 $\| G(x) \|_2 = \| \nabla^2 f(x) \|_2 \leqslant M, x \in L(x^{(0)})$.由 $\{x^{(k)}\} \subset L(x^{(0)})$ 及 $y_k = [\int_0^1 \nabla^2 f(x^{(k)} + ts_k)\mathrm{d}t]s_k$,知 $\frac{\| y_k \|_2}{\| s_k \|_2} \leqslant M, \frac{\| y_k \|_2^2}{s_k^T y_k} \leqslant M$.

$k = 0$ 时,$d^{(0)} = -g^{(0)}, -g^{(0)^T} d^{(0)} = \| g^{(0)} \|_2^2 \geqslant \frac{1}{M} \| g^{(0)} \|_2^2$.假设 $k = t$ 时,有 $-g^{(t)^T} d^{(t)} \geqslant \frac{1}{M} \| g^{(t)} \|_2^2$,则 $k = t + 1$ 时.

若 $s_t^T y_t = \| s_t \|_2 \| y_t \|_2$,则 $d^{(t+1)} = -\frac{\| s_t \|_2}{\| y_t \|_2} g^{(t+1)}, -g^{(t+1)^T} d^{(t+1)} = \frac{\| s_t \|_2}{\| y_t \|_2} \| g^{(t+1)} \|_2^2 \geqslant \frac{1}{M} \| g^{(t+1)} \|_2^2$,若 $s_t^T y_t < \| s_t \|_2 \| y_t \|_2$,则由 wolfe 线搜索准则知

$$s_t^T y_t \geqslant (1 - \sigma)\alpha_t(-g^{(t)^T} d^{(t)}) \geqslant \frac{1}{M}(1 - \sigma)\alpha_t \| g^{(t)} \|_2^2 > 0.$$

其中 $0 < \mu < \sigma < 1$.设 s_t 与 y_t 夹角为 $\theta_t = 4\alpha > 0$,显然

$$l^{(t)} = \frac{[\| y_t \|_2 s_t + \| s_t \|_2 y_t]}{\| \| y_t \|_2 s_t + \| s_t \|_2 y_t \|_2} \in \mathrm{span}\{s_t, y_t\}$$

是 θ_t 的平分线.令 $l_1 = l^{(t)}, l_2 \in \mathrm{span}\{s_t, y_t\}$ 且 $l_2 \perp l_1$ 及 $\| l_2 \|_2 = 1$.注意到 $\| l_1 \|_2 = \| l_2 \|_2 = 1, l_1 \perp l_2, s_t, y_t \in \mathrm{span}\{l_1, l_2\}$,不妨设 $s_t = \| s_t \|_2(l_1\cos 2\alpha + l_2 \sin 2\alpha), y_t = \| y_t \|_2(l_1 \cos 2\alpha - l_2 \sin 2\alpha)$.
则

$$u^{(t)} = \frac{(1 - \cos 2\alpha) l_1 - (\sin 2\alpha) l_2}{\sqrt{(1 - \cos 2\alpha)^2 + \sin^2 2\alpha}} = l_1 \sin\alpha - l_2 \cos\alpha,$$

$$v^{(t)} = \frac{(1 - \cos 2\alpha) l_1 + (\sin 2\alpha) l_2}{\sqrt{(1 - \cos 2\alpha)^2 + \sin^2 2\alpha}} = l_1 \sin\alpha + l_2 \cos\alpha,$$

$$u^{(t)^T} v^{(t)} = \sin^2\alpha - \cos^2\alpha = -\cos 2\alpha,$$

将 $g^{(t+1)}$ 分解为 $g^{(t+1)} = g_1 + g_2$,其中 $g_2 \in \mathrm{span}\{s_t, y_t\}, g_1 \in \mathrm{span}\{s_t, y_t\}^{\perp}$.
设 $g_2 = \| g_2 \|_2(l_1 \cos\beta + l_2 \sin\beta)$,则

$$u^{(t)^T} g_2 = \| g_2 \|_2(\sin\alpha\cos\beta - \cos\alpha\sin\beta) = \| g_2 \|_2 \sin(\alpha - \beta)$$

$$v^{(t)^T} g_2 = \| g_2 \|_2(\sin\alpha\cos\beta + \cos\alpha\sin\beta) = \| g \|_2 \sin(\alpha + \beta)$$

$$\begin{aligned} g_2^T H_1^{(t)} H_2^{(t)} g_2 &= \| g_2 \|_2^2 - 2(u^{(t)^T} g_2)^2 - 2(v^{(t)^T} g_2)^2 + 4(u^{(t)^T} g_2)(v^{(t)^T} g_2)(u^{(t)^T} v^{(t)}) \\ &= \| g_2 \|_2^2 [1 - 2\sin^2(\alpha - \beta) - 2\sin^2(\alpha + \beta) - 4\sin(\alpha - \beta)\sin(\alpha + \beta)\cos 2\alpha] \\ &= \| g_2 \|_2^2 (2\cos^2 2\alpha - 1) = \| g_2 \|_2^2 \cos\theta_t \end{aligned}$$

由于 $g_1 \perp u^{(t)}, g_1 \perp v^{(t)}$,故 $H_1^{(t)} H_2^{(t)} g_1 = g_1$,因此由 $0 < \frac{\| y_t \|_2^2}{s_t^T y_t} \leqslant M$ 知

$$\begin{aligned} -g^{(t+1)^T} d^{(t+1)} &= \frac{\| s_t \|_2}{\| y_t \|_2}(g_1 + g_2)^T H_1^{(t)} H_2^{(t)}(g_1 + g_2) = \frac{\| s_t \|_2}{\| y_t \|_2}(\| g_1 \|_2^2 + \| g_2 \|_2^2 \cos\theta_t) \\ &\geqslant \frac{\| s_t \|_2}{\| y_t \|_2} \| g^{(t+1)} \|_2^2 \cos\theta_t = \| g^{(t+1)} \|_2^2 \frac{\| s_t \|_2}{\| y_t \|_2} \frac{s_t^T y_t}{\| s_t \|_2 \| y_t \|_2} \\ &= \| g^{(t+1)} \|_2^2 \frac{s_t^T y_t}{\| y_t \|_2^2} \geqslant \frac{1}{M} \| g^{(t+1)} \|_2^2. \end{aligned}$$

定理 2 在定理 1 的条件下,由算法 1 产生的 $\{x^k\}$ 的任一极限点均是 $f(x)$ 的整体最优解.

证明 由 wolfe 线搜索准则知

$$f(x^{(k)}) - f(x^{(k+1)}) \geqslant \alpha_k \mu(-g^{(k)T} d^{(k)}), \qquad s_k^T y_k \geqslant (1 - \sigma)\alpha_k(-g^{(k)^T} d^{(k)}).$$

由 $d^{(k)} = -\frac{\| s_{k-1} \|_2}{\| y_{k-1} \|_2} g^{(k)}$ 或 $d^{(k)} = -\frac{\| s_{k-1} \|_2}{\| y_{k-1} \|_2} H_1^{(k-1)} H_2^{(k-1)} g^{(k)}$ 知 $\| d^{(k)} \|_2^2 = d^{(k)^T} d^{(k)} = \frac{\| s_{k-1} \|_2^2}{\| y_{k-2} \|_2^2} \| g^{(k)} \|_2^2$.

由定理1知

$$(-g^{(k)^T}d^{(k)}) \geqslant \frac{\| s_{k-1}\|_2}{\| y_{k-1}\|_2}\| g^{(k)}\|_2^2\cos\theta_{k-1} \geqslant \frac{1}{M}\| g^{(k)}\|_2^2.$$

下面先证明$\liminf\limits_{k\to\infty}\| g^{(k)}\|_2 = 0$.若不然,则存在 $\varepsilon > 0$,使对 $\forall k$ 有$\| g^{(k)}\|_2^2 \geqslant \varepsilon$,于是

$$f(x^{(k)}) - f(x^{(k+1)}) \geqslant \alpha_k\mu(-g^{(k)^T}d^{(k)}) \geqslant \frac{1}{M}\alpha_k\mu\| g^{(k)}\|_2^2 \geqslant \frac{1}{M}\alpha_k\mu\varepsilon.$$

由 $L(x^{(0)})$ 有界,$f(x)$ 二阶连续可微,$\{f(x^{(k)})\}$ 单调减,可知存在$f^* > -\infty$ 使$f(x^{(R)}) \to f^*$,故

$$+\infty > f(x^{(0)}) - f^* = \sum_{k=0}^{\infty}[f(x^{(k)}) - f(x^{(k+1)})] \geqslant \frac{1}{M}\mu\varepsilon\sum_{k=0}^{\infty}\alpha_k.$$

因而 $k\to+\infty$ 时,$\alpha_k\to 0$.于是存在 $k_0 > 0$,当 $k \geqslant k_0$ 时,有$\frac{\alpha_k}{1-\sigma} < \frac{1}{2}$.由于

$$\frac{\| s_k\|_2^2}{s_k^Ty_k} \leqslant \frac{\alpha_k^2\| d^{(k)}\|_2^2}{(1-\sigma)\alpha_k(-g^{(k)^T}d^{(k)})} \leqslant \frac{\dfrac{\alpha_k\| s_{k-1}\|_2^2\| g^{(k)}\|_2^2}{\| y_{k-1}\|_2^2}}{\dfrac{(1-\sigma)\| s_{k-1}\|_2\| g^{(k)}\|_2^2\cos\theta_k}{\| y_{k-1}\|_2}} = \left(\frac{\alpha_k}{1-\sigma}\right)\frac{\| s_{k-1}\|_2^2}{s_{k-1}^Ty_{k-1}}.$$

可知 $k > k_0$ 时有

$$\frac{\| s_k\|_2^2}{s_k^Ty_k} \leqslant \left(\prod_{i=k_0+1}^{k}\frac{\alpha_i}{1-\sigma}\right)\frac{\| s_{k_0}\|_2^2}{s_{k_0}^Ty_{k_0}} \leqslant \left(\frac{1}{2}\right)^{k-k_0}\cdot\frac{\| s_{k_0}\|_2^2}{s_{k_0}^Ty_{k_0}}.$$

因而 $k\to+\infty$ 时,$\frac{\| s_k\|_2^2}{s_k^Ty_k}\to 0$.

但由 $L(x^{(0)})$ 有界,$f(x)$ 为二阶连续可微凸函数,$y_k = [\int_0^1\nabla^2 f(x^{(k)} + ts_k)\mathrm{d}t]s_k$,知存在 $M' > 0$,使$\frac{s_k^Ty_k}{\| s_k\|_2^2} \leqslant M'$,再由 $s_k^Ty_k \geqslant (1-\sigma)\alpha_k(-g^{(k)^T}d^{(k)}) > 0$,知$\frac{\| s_k\|_2^2}{s_k^Ty_k} \geqslant \frac{1}{M'} > 0$.此与$\frac{\| s_k\|_2^2}{s_k^Ty_k}\to 0$ 矛盾.由此得$\liminf\limits_{k\to\infty}\| g^{(k)}\|_2 = 0$.

下面再证$\{x^{(k)}\}$ 的任一极限点 $\bar{x}$ 均为$f(x)$ 的整体最优解.设$\{x^{(k_i)}\}\subset\{x^{(k)}\}$ 使 $x^{(k_i)}\to\bar{x}$,显然$f(\bar{x}) = f^*$.由于$\liminf\limits_{k\to\infty}\| g^{(k)}\|_2 = 0$,因此$\{x^{(k)}\}$ 有一个极限点 $\hat{x}$,使 $g(\hat{x}) = 0$,$f(\hat{x}) = f^*$.由 $f(x)$ 可微凸知 $\hat{x}$ 为$f(x)$ 的整体最优解,即f^* 为$f(x)$ 的整体最优值.再由 $f(\bar{x}) = f^*$,知 $\bar{x}$ 也为$f(x)$ 整体最优解.

[参考文献]

[1] 尉继英.一种求解无约束极值问题的无记忆拟牛顿算法[J].计算数学,1990,11(3):259—269.
[2] 徐成贤,陈志平,李乃成.近代优化方法[M].北京:科学出版社,2002.

The Convergence of a Memoryless Quasi-Newton Method

Yan Shijian

(School of Mathematics and Computer Science, Nanjing Normal University, 210097, Nanjing, China)

Abstract: Under the condition that $f(x)$ is a twice continuously differentiable convex function, the convergence of a memoryless Quasi-Newton method is proved.

Key words: convex function, memoryless Quasi-Newton method, convergence

[责任编辑:陆炳新]