广西科学 Guangxi Sciences 2006，13(3)：187～189

一个非单调 BFGS 信赖域算法

A Nonmonotone BFGS-Trust-Region Algorithm

吴庆军
WU Qing-jun

(玉林师范学院数学与计算机科学系，广西玉林　537000)
(Department of Mathematics and Computer Science, Yulin Teachers's College, Yulin, Guangxi, 537000, China)

摘要：将新的 BFGS 校正公式 $B_{k+1}=B_k+\frac{y_k^* y_k^{*T}}{s_k^T y_k^*}-\frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}$，与文献[16]中的算法相结合给出一个非单调 BFGS 校正的信赖域算法. 该算法在假设条件：(i)存在常数 c_1,c_2,c_3，使得对所有的 $\Delta_k>0, g_k\in R^n$，对称正定阵 $B_k\in R^{n\times n}$，有 $pred_k\geqslant c_1\|g_k\|\min\{\Delta_k,c_2\|g_k\|,c_3\|g_k\|/\|B_k\|\}$；(ii)若 $\|B_k^{-1}\|\leqslant\Delta_k$，则 $d_k=-B_k^{-1}g_k$；(iii)$f(x)$ 是二次连续可微函数，$\nabla^2 f(x_k)$ 是 *Lipschitz* 连续，水平集 $\phi(x_0)$ 有界下，具有全局收敛性和 Q- 二次收敛性.

关键词：非单调　BFGS 校正　全局收敛性　信赖域算法
中图法分类号：O221　　**文献标识码**：A　　**文章编号**：1005-9164(2006)03-0187-03

Abstract: A new nonmonotone BFGS-trust-region algorithm is proposed by combining the BFGS update $B_{k+1}=B_k+\frac{y_k^* y_k^{*T}}{s_k^T y_k^*}-\frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}$, with the algorithm given in references [16]. The global and Q-quadratic convergences of the proposed algorithm are also proved under the following conditions: (i) there are constants c_1,c_2,c_3 such that $pred_k\geqslant c_1\|g_k\|\min\{\Delta_k,c_2\|g_k\|,c_3\|g_k\|/\|B_k\|\}$ for $\Delta_k>0, g_k\in R^n$ and some symmetry positive definite matrix $B_k\in R^{n\times n}$; (ii) $d_k=-B_k^{-1}g_k$ if $\|B_k^{-1}g_k\|\leqslant\Delta_k$ and (iii) $f(x)$ is a second-order continuously differentiable function, $\nabla^2 f(x)$ is Lipschitz continuous and the level set $\phi(x_0)$ is bounded.

Key words: nonmonotone, BFGS update, global convergence, trust-region algorithm

无约束最优化问题

$$\min_{x\in R^n} f(x), \tag{1}$$

其中 $f(x)$ 是连续可微函数，BFGS 方法是求解无约束最优化的有效的拟牛顿方法[1~3]，文献[4～7]也给出了一些修改的 BFGS 方法并分析了其相应算法的收敛性. 一个大家都知道的、重要的 BFGS 校正公式是：

$$B_{k+1}=B_k+\frac{y_k y_k^T}{s_k^T y_k}-\frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \tag{2}$$

其中 $s_k=x_{k+1}-x_k, y_k=g_{k+1}-g_k, g_k$ 和 g_{k+1} 分别是 $f(x)$ 在 x_k 和 x_{k+1} 处的梯度值. 在文献[8]中也给出了一个新的 BFGS 校正公式

$$B_{k+1}=B_k+\frac{y_k^* y_k^{*T}}{s_k^T y_k^*}-\frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \tag{3}$$

其中 $y_k^*=\frac{y_k^T s_k}{|y_k^T s_k|}y_k$. BFGS 信赖域算法是非常有效的算法，文献[9～13]分析了信赖域算法的收敛性，但是，这些算法均是单调算法. 文献[16]提出了一类非单调信赖域算法，其算法在某一步后，每试探一步要求当前点的函数值与前不固定个点(如 M_{k+1})中函数值最大的进行比较，而且 M_k 可进行调整，对目标函数，当性态好时，M_k 减少甚至为零，即算法采用单调下降步；当性态坏时，M_k 愈大，算法采用非单调下降步.

本文将公式(3)与文献[16]中的算法相结合给出一个新的算法，此算法在较少的假设条件下同样具有全局收敛性和 Q- 二次收敛性. 本文采用下列记号：$\|\cdot\|$ 是 $R1$ 中的 Euclid 范数，$[b_1,\cdots,b_n]$ 是由向量 $b_1,\cdots,b_n$ 生成的 R^n 子空间，$g(x)\in R^n$ 是 f 在 x 处的

收稿日期：2006-02-23
修回日期：2006-05-24
作者简介：吴庆军(1962-)，男，广西桂平人，讲师，主要从事运筹学研究。

梯度值,$\{x_n\}$ 是由算法产生的点列,记 $f_k := f(x_k)$, $g_k := \nabla f(x_k)$. 以下部分讨论的 B_k 都是在 B_0 为对称正定时产生的.

1 算法及其性质

首先给出解问题(1)的一般信赖域算法子问题的定义式:

$$\min \psi_k(d) = f_k + g_k^T d + \frac{1}{2} d^T B_k d, d \in R^n, \quad (4)$$

$$\text{s.t.} \ \| d \| \leqslant \Delta_k,$$

其中,$\Delta_k > 0$ 是信赖域半径,d 是试探步,B_k 是 $f(x)$ 在 x_k 的 Hessian 矩阵或其近似. 文献[14,15]给出求解信赖域子问题的一个算法并讨论其收敛性. 定义

$$f_{m(k)} = \max_{0 \leqslant j \leqslant l(k)} \{f_{k-j}\}, \quad (5)$$

其中,$l(k) = \min\{l(k-1)+1, Mk\}$, Mk 是非负整数,$l(0) = 0$, 下面给出一个非单调BFGS校正的信赖域算法.

算法 算法的步骤如下:

步骤 0:给定 $x_0 \in R^n$, 对称矩阵 $B_0 \in R^{n\times n}$, $\Delta_0 > 0$, $0 < \mu < 1$, $0 < \gamma < 1$, $\eta > 0$, $c \geqslant 1$, $\varepsilon > 0$, 整数 $M \geqslant 0$, $M_0 := M$, $l(0) := 0$, $k := 0$.

步骤 1:计算 g_k. 如果 $\| g_k \| < \varepsilon$, 则终止;否则转步骤 2.

步骤 2:解子问题(4),得试探步 d_k.

步骤 3:计算 $\rho^k m(k) = \dfrac{ared^k_{m(k)}}{pred_k}$ 和 $u_k = \min\{u, \eta\Delta_k \dfrac{\| g_k \|}{pred_k}\}$, 其中, $ared^k_{m(k)} = f_{m(k)} - f(x_k + d_k)$, $pred_k = - g_k^T d_k - \dfrac{1}{2} d_k^T B_k d_k$.

步骤 4:如果 $\rho^k m(k) < u_k$, 则取新信赖域半径 $\Delta_k^{new} := \gamma\Delta_k$, $M_k^{new} := M_k$, 转步骤 2(称为内循环);否则令 $x_{k+1} = x_k + d_k$(称为成功步), $\Delta_{k+1} \geqslant \Delta_k$, 如果 $\rho^k m(k) > c$, 则 $M_{k+1} := 0$, 否则 $M_{k+1} := M$, 利用校正公式(3)产生 B_{k+1}, 令 $k := k+1$, 转步骤 1.

当 $M = 0$ 时,算法成为一般单调信赖域算法. 当 $y_k^T s_k > 0$ 时,BFGS 校正公式保持校正矩阵的正定性. B_k 正定时,由

$$y_k^{*T} s_k = \frac{y_k^T s_k}{|y_k^T s_k|} y_k^T s_k = \frac{(y_k^T s_k)^2}{|y_k^T s_k|} = |s_k^T B_k s_k + o(\| s_k \|^2)| \geqslant |\lambda_k + o(1)\| |s_k\|^2 > 0, \quad (6)$$

其中 λ_k 是 B_k 的最小特征值,$s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$. 所以 B_{k+1} 是正定矩阵. 当 B_0 取为正定矩阵时,可用数学归纳法证得由校正公式(3)产生的 B_k 都是正定的. 综上所述,由 B_k 是正定的,则(4)式是一个具有二次约束的严格二次规划,且若 $\lim_{k\to\infty} x_k = x^*$, 则得到的 $\nabla^2 f(x^*)$ 是正定的.

2 收敛性分析

为了证明算法的收敛性,首先给出下面的假设条件:

(i) 存在常数 c_1, c_2, c_3, 使得对所有的 $\Delta_k > 0$, $g_k \in R^n$, 对称正定阵 $B_k \in R^{n\times n}$, 有

$$pred_k \geqslant c_1 \| g_k \| \min\{\Delta_k, c_2 \| g_k \|, c_3 \| g_k \| / \| B_k \|\};$$

(ii) 若 $\| B_k^{-1} g_k \| \leqslant \Delta_k$, 则 $d_k = - B_k^{-1} g_k$.

首先,可以证明算法是适定的(其定义和证明见文献[16]).

定理 1 若(i)成立,则算法是适定的.

定理 1 说明每一步迭代的内循环必在有限步终止,否则成功步数无穷,它产生无穷序列 $\{x_k\}$. 定义水平集 $\phi(x_0) = \{x \in R^n: f(x) \leqslant f(x_0)\}$. 由文献[16]知下面的两个引理成立.

引理 1 若(i)成立,$\{x_k\} \subset \phi(x_0)$.

引理 2 若(i)成立,如果 $\{f_k\}$ 有下界,则 $\{f_{l(k)}\}$ 非增且收敛.

定理 2 假设水平集 $\phi(x_0)$ 有界且(i)成立. 如果存在正常数 ε, 使得对所有的 k, $\| B_k \| \leqslant \varepsilon$, 则 $\liminf_{k\to\infty} \| g_k \| = 0$.

证明 假设结论不成立,则存在正常数 ε_1, 使得对所有的 k, $\| g_k \| \geqslant \varepsilon_1$. 容易证得

$$pred_k \geqslant c_1\varepsilon_1 \min\{\Delta_k, c_2\varepsilon_1, c_3\varepsilon_1/\varepsilon\} > 0, \forall\ k, \quad (7)$$

$$f_{m(k)} - f_{k+1} \geqslant \min\{\eta\varepsilon_1\Delta_k, \mu c_1\varepsilon_1 \min\{\Delta_k, c_2\varepsilon_1, c_3\varepsilon_1/\varepsilon\}\}, \quad (8)$$

所以 $\lim_{k\to\infty} \Delta_{m(k)-1} = 0$. 可以证明 $\lim_{k\to\infty} \Delta_k = 0$. 事实上,令 $m'(k) = m(k+M+2)$, 因此 $m'(k) \geqslant k + M + 2 - M = k + 2$, 利用归纳法可以证明得

$$\lim_{k\to\infty} \Delta_{m'(k)-j} = 0. \lim_{k\to\infty} f_{m'(k)-j} = \lim_{k\to\infty} f_{m(k)}, \forall\ j \geqslant 1. \quad (9)$$

所以

$$x_{k+1} = x_{m'(k)} - \sum_{j=1}^{m'(k)-(k+1)} d_{m'(k)-j}. \quad (10)$$

由 $m'(k)$ 的定义,得 $m'(k) - (k+1) = m(k+2+m) - (k+1) \leqslant M+1$, 于是(9)式和(10)式隐含 $\lim_{k\to\infty} \| x_{k+1} - x_{m(k)} \| = 0$. 由 $\psi(x_0)$ 的有界性可知, $\{f_k\}$ 有下界,从而

$$\lim_{k\to\infty} f_{k+1} = \lim_{k\to\infty} f_{m'(k)} = \lim_{k\to\infty} f_{m(k)}, \forall\ j \geqslant 1. \quad (11)$$

因此(8)式和(11)式隐含 $\lim_{k\to\infty} \Delta_k = 0$. 可以证明

$$\rho^k_{m(k)} = \frac{f_{m(k)} - f_{k+1}}{pred_k} \geqslant \frac{f_k - f_{k+1}}{pred_k}, \forall\ k$$

$$\left|\frac{f_k - f_{k+1}}{pred_k} - 1\right| \leqslant$$

$$\frac{\| \nabla f(x_k + \theta_k d_k) - g_k \| \| d_k \| + \frac{1}{2}\varepsilon \| d_k \|^2}{c_1 \varepsilon_1 \min\{\Delta_k, c_2\varepsilon_1, c_3\varepsilon_1/\varepsilon\}}, \forall k \tag{12}$$

其中 $\theta_k \in (0,1)$. 令 $k \to \infty$, 有 $\lim_{k\to\infty} \frac{f_k - f_{k+1}}{pred_k} = 1$. 由(12)式可得对充分大的 k, $\frac{f_{m(k)} - f_{k+1}}{pred_k} \geqslant \mu \geqslant \mu_k$, 因此由算法得, 对充分大的 k, $\Delta_{k+1} \geqslant \Delta_k$, 与 $\lim_{k\to\infty}\Delta_k = 0$ 矛盾. 所以命题得证.

如果条件"$g_k = \nabla f(x_k)$"改为"g_k 是 $\nabla f(x_k)$ 的近似", 相应的定理1,2和引理1,2的结论也成立. 为证明算法的 Q- 二次收敛性, 需要下面的假设条件.

(iii) $f(x)$ 是二次连续可微函数, $\nabla^2 f(x_k)$ 是 Lipschitz 连续, 水平集 $\phi(x_0)$ 有界.

定理 3 对所有 k, $B_k = \nabla^2 f(x_k)$. 如果(i), (ii) 和(iii)成立, 则 $\{x_k\}$ Q- 二次收敛于 x^*.

证明 由 $B_k = \nabla^2 f(x_k)$ 是正定的, $x_k \to x^*$, $\nabla^2 f(x^*)$ 也是正定的, 令 $\kappa = \{k: \| B_k^{-1} g_k \| > \Delta_k\}$. 假设集合 κ 有无穷多个元素. 由(ii)得

$$\| g_k \| \geqslant \frac{\Delta_k}{\| B_k^{-1} \|}, \forall k \in \kappa; \| g_k \| \geqslant \frac{\| d_k \|}{B_k^{-1}}, \forall k$$

可以证明

$$pred_k \geqslant \frac{c_1 \| \Delta_k^2 \|}{2 \| B_*^{-1} \|} \min\{1, \frac{c_2}{\| B_*^{-1} \|}, \frac{c_3}{\| B_*^{-1} \| \| B_* \|}\} > 0, \forall k \in \kappa$$

$$pred_k \geqslant \frac{c_1 \| d_k^2 \|}{2 \| B_*^{-1} \|} \min\{1, \frac{c_2}{\| B_*^{-1} \|}, \frac{c_3}{\| B_*^{-1} \| \| B_* \|}\} > 0, \forall k$$

其中, $B_* = \nabla^2 f(x^*)$, $B_*^{-1} = (\nabla^2 f(x^*))^{-1}$. 从而

$$\lim_{k(\in)\to\infty} \Delta_k = 0, \lim_{k\to\infty} \| d_k \| = 0. \tag{13}$$

易证

$$\left|\frac{f_k - f_{k+1}}{pred_k} - 1\right| \leqslant \frac{L \| d_k \| \| B_*^{-1} \|}{c_1 \min\{1, \frac{c_2}{\| B_*^{-1} \|}, \frac{c_3}{\| B_*^{-1} \| \| B_* \|}\}}$$

其中 L 是 Lipschitz 常数. 因此对充分大的 k, $(f_{m(k)} - f_{k+1})/pred_k \geqslant \mu_k$. 于是对充分大的 k, $\Delta_{k+1} \geqslant \Delta_k$ 与(13)式的第一式矛盾. 因此对充分大的 k, $\| B_k^{-1} g_k \| \leqslant \Delta_k$, 由假设(ii)可得, $d_k = -B_k^{-1} g_k$, 这说明 $\{x_k\}$ Q- 二次收敛 x^*.

参考文献:

[1] DENNIS J E, SCHNABEL R B. Numerical methods for unconstrained optimization and nonlinear equations[M]. NJ: Pretice-Hall Inc Englewood Chiffs, 1983.

[2] FLETCHER R, WILEY JOHN, SONS. Practical methods of optimization[M]. Chichester: Wiley & Sons, 1987.

[3] DENNIS J E J R, MORE J J. A characterization of superlinear convergence and its application to quasi-Newton methods[J]. Math Comp, 1974, 28: 1171-1190.

[4] LI D, FUKUSHIMA M. A modified BFGS method and its global convergence in nonconvex minimization[J]. Journal of Computational and Applied Mathematics, 2001, 129: 15-35.

[5] POWELL M J D. A new algorithm for unconstrained optimization[M]//ROSEN J B, MANGASARIAN O L, RITTER K, eds. Nonlinear programming. New York: Academic Press, 1970.

[6] WEI Z, QI L, CHEN X. An SQP-type method and its application in stochastic programming[J]. Journal of Optimization Theory and Applications, 2003, 116: 205-228.

[7] WEI Z, YU G, YUAN G, et al. The superlinear convergence of a modified BFGS-type method for unconstrained optimization[M]. Computational and Applications, 2002.

[8] 袁功林, 韦增欣. 一个新的 BFGS 信赖域算法[J]. 广西科学, 2004, 11(3): 195-196.

[9] BACHEM A, GROTSCHEL M, KORTE B. Mathematical programming: the state of Art[M]. Berlin: Springer-VERlag, 1983: 258-287.

[10] SHULTZ G A, SSHNABEL R B, BYRD R H. A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties[J]. SIAM J Number Anal, 1985, 22: 47-67.

[11] 袁亚湘. 信赖域方法的收敛性[J]. 计算数学, 1994, 16(3): 333-346.

[12] BULEAU J P, VIAL J Ph. Curvilinear path and trust region in unconstrained optimization, a convergence analysis[J]. Math Prog Study, 1987, 30: 82-101.

[13] BYRD B H, SCHNABEL R B, SHULTZ G A. Approximate solution of the trust region problem by minimization over two-dimensional subspaces[J]. Math Prog, 1988, 40: 247-263.

[14] NOCEDAL J, YUAN Y Y. Combining trust-region and line search techniques[M]//Yuan Y Y. Advance in nonlinear programming. Applied Optimization, Kluwer Academic Publ, Dordrecht, 1998: 153-157.

[15] RENDL F, WOLKOWICZ H. A semidefinite framework for trust region subproblems with applications to large scale minimization[J]. Math Program, 1997, 77: 273-299.

[16] 柯小五, 韩继业. 无约束最优化的一类非单调信赖域算法[J]. 中国科学, 1998, 28(6): 488-492.

(责任编辑: 邓大玉 凌汉恩)