

文章编号: 1007- 2985(2009) 06- 0032- 03

一类非单调自适应- BFGS 信赖域算法^{*}

景书杰, 张小亮

(河南理工大学数学与信息科学学院, 河南 焦作 454003)

摘 要: 针对无约束优化问题, 每次迭代充分利用当前迭代点及其一阶导数的信息自动生成一个信赖域半径, 并结合 BFGS 算法的优点, 构造了无约束优化问题的一种非单调自适应- BFGS 算法. 在一定条件下, 给出了算法的全局收敛性以及具有超线性收敛速度的证明.

关键词: 无约束优化; 非单调算法; 自适应信赖域算法; BFGS 校正; 全局收敛性
中图分类号: O244 文献标识码: A

信赖域算法是一类重要的数值计算方法, 由于具有较好的可靠性和很强的收敛性, 因而在近 30 a 来受到最优化研究界的重视. 特别是近几年来, 一直是最优化领域中的一个研究热点.

无约束优化问题

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

其中 $f(x): \mathbf{R}^n \rightarrow \mathbf{R}$ 是连续可微函数. 信赖域方法是一种迭代的方法, 每次迭代要求计算信赖域子问题

$$\min_k g_k^T d + \frac{1}{2} d^T B_k d \quad \text{s. t.} \quad \|d\| \leq \rho_k. \tag{2}$$

其中: $g_k = -\nabla f(x_k)$; B_0 为对称正定矩阵; B_k 是由 BFGS 校正公式产生的对称正定矩阵; ρ_k 是信赖域半径. 文献[1] 给出了一个自动确定初始信赖域半径的 ITRR 算法, 文献[2- 3] 给出了一些自适应信赖域半径的选取机制, 具体如下:

- (1) $\rho_k = \frac{c^p \|g_k\|}{\|g_k\| B_k}, 0 < c < 1, \rho_k = \min\{\rho_k, 1\}$, 其中 p 是非负整数.
- (2) $\rho_k = \frac{c^p \|g_k\|^3}{g_k^T B_k g_k}, 0 < c < 1$, 其中 p 是非负整数.

BFGS 校正公式是非常好的一种拟牛顿方法, 利用 BFGS 公式可以得到对称正定的校正矩阵 B_k , 从而保证算法的总体收敛性. BFGS 校正公式为

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

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.

假设 $y_k^* = \frac{y_k^T s_k}{|y_k^T s_k|} y_k$, 得到另一种新的 BFGS 校正公式^[8]:

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

笔者将非单调信赖域算法^[4-6]、自适应算法^[1-3, 6-11] 以及 BFGS 校正方法^[11-16] 相结合应用到无约束非单调优化问题, 构造了无约束优化问题的一种具有超线性收敛速度的非单调自适应- BFGS 算法.

^{*} 收稿日期: 2009- 07- 22
基金项目: 国家自然科学基金资助项目(10671057)
作者简介: 景书杰(1965-), 男, 河南长垣人, 河南理工大学数学与信息科学学院副教授, 博士, 主要从事最优化理论及应用研究.

1 非单调自适应- BFGS 信赖域算法

通过解决信赖域的子问题可以得到试探步 $\mathbf{d}_k$, 同时定义 $f_{l(k)} = f(\mathbf{x}_{l(k)}) = \max_{0 \leq j \leq m(k)} f(\mathbf{x}_{k-j})$, 其中 $m(0) = 0, 0 \leq m(k) \leq \min\{m(k-1) + 1, M\}, k = 1, M$ 是非负整数. 定义实际下降量为 $\text{Are } \mathbf{d}_k = f(\mathbf{x}_{l(k)}) - f(\mathbf{x}_k + \mathbf{d}_k)$, 预测下降量为 $\text{Pre } \mathbf{d}_k = \mathbf{g}_k^T \mathbf{d}_k - \frac{1}{2} \mathbf{d}_k^T \mathbf{B}_k \mathbf{d}_k$. 现给出如下算法:

- 步骤 0 给定 $\mathbf{x}_0 \in \mathbf{R}^n$, 对称矩阵 $\mathbf{B}_0 \in \mathbf{R}^{n \times n}, \mathbf{B}_0 > 0, 0 < u < 1, 0 < h < 1, 0 < \gamma < \frac{1}{2}$, 整数 $M \geq 0, m(0) = 0, k = 0, i = 0$;
- 步骤 1 计算 $\mathbf{g}_k$, 若 $\|\mathbf{g}_k\| \leq \epsilon$, 则终止; 否则转步骤 2;
- 步骤 2 取 $\mathbf{d}_k = h^i \min\{\frac{\mathbf{d}_{k-1}}{\|\mathbf{y}_k\|}, \frac{\mathbf{g}_k}{\|\mathbf{g}_k\|}\}$, 解信赖域子问题, 得到试探步 $\mathbf{d}_k$;
- 步骤 3 计算 $r_k = \frac{\text{Are } \mathbf{d}_k}{\text{Pre } \mathbf{d}_k}$ 和 $u_k = \min\{u, \gamma - \frac{\mathbf{g}_k^T \mathbf{d}_k}{\text{Pre } \mathbf{d}_k}\}$;
- 步骤 4 若 $r_k \geq u_k$, 则令 $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k, m(k+1) = \min\{m(k) + 1, M\}$; 否则, 令 $i = i + 1$ 转步骤 1;
- 步骤 5 利用新的 BFGS 校正公式 $\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{s}_k^T \mathbf{y}_k} - \frac{\mathbf{B}_k \mathbf{s}_k \mathbf{s}_k^T \mathbf{B}_k}{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k}$, 产生 $\mathbf{B}_{k+1}$, 令 $i = 0, k = k + 1$, 转步骤 1.

2 算法的收敛性分析

假设: (1) 对任意的 k , 存在有界闭集 S_k 使得 $\mathbf{x}_k, \mathbf{x}_{k+1} \in S_k$; (2) $f(\mathbf{x})$ 二阶连续可微, $f(\mathbf{x})$ 一致连续, $\nabla^2 f(\mathbf{x})$ 和 $\mathbf{B}_k$ 是一致有界的.

- 引理 1^[9] $\text{Are } \mathbf{d}_k - \text{Pre } \mathbf{d}_k = o(\|\mathbf{d}_k\|^2)$.
- 引理 2^[11] 设 $\mathbf{d}_k$ 是子问题(2) 的解, 则有

$$\text{Pre } \mathbf{d}_k = \mathbf{g}_k^T \mathbf{d}_k - \frac{1}{2} \mathbf{d}_k^T \mathbf{B}_k \mathbf{d}_k \leq \mathbf{g}_k^T \min\{\frac{\mathbf{d}_{k(i)}}{\|\mathbf{d}_{k(i)}\|}, \frac{\mathbf{g}_k}{\|\mathbf{g}_k\|}\}.$$

定理 1 算法每一步迭代的内循环必在有限步终止, 否则成功步数无穷, 产生无穷序列 $\{\mathbf{x}_k\}$, 其中定义水平集为 $\mathcal{L}(\mathbf{x}_0) = \{\mathbf{x} \in \mathbf{R}^n : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$.

证明 假设结论不成立, 即算法中步骤 1 和步骤 4 之间的内循环不有限终止. 令 $k(i)$ 为 $\mathbf{x}_k$ 处第 i 次内循环, 则 $r_{k(i)} \geq u_k, i = 1, 2, \dots$, 而且 $\|\mathbf{d}_{k(i)}\| \rightarrow 0, i \rightarrow \infty$. 又由引理 1 和引理 2, 得

$$|r_{k(i)} - 1| = |\frac{\text{Are } \mathbf{d}_{k(i)} - \text{Pre } \mathbf{d}_{k(i)}}{\text{Pre } \mathbf{d}_{k(i)}}| = \frac{o(\|\mathbf{d}_{k(i)}\|^2)}{\mathbf{g}_k^T \mathbf{d}_{k(i)} - \frac{1}{2} \mathbf{d}_{k(i)}^T \mathbf{B}_k \mathbf{d}_{k(i)}} \leq \frac{o(\|\mathbf{d}_{k(i)}\|^2)}{\frac{1}{2} \mathbf{g}_k^T \min\{\frac{\mathbf{d}_{k(i)}}{\|\mathbf{d}_{k(i)}\|}, \frac{\mathbf{g}_k}{\|\mathbf{g}_k\|}\}} \rightarrow 0.$$

故当 $i \rightarrow \infty$ 时, $|r_{k(i)} - 1| \rightarrow 0$, 即 $r_{k(i)} \rightarrow 1$, 与 $r_{k(i)} \geq u_k < 1$ 矛盾. 所以假设不成立, 原定理成立.

- 引理 3^[6] 若假设条件(2) 成立, 且 $\{f_k\}$ 有下界, 则数列 $\{f_{l(k)}\}$ 是单调非增且收敛的.
- 定理 2 假设水平集 $\mathcal{L}(\mathbf{x}_0)$ 有界且假设条件(1), (2) 成立, 且存在正常数 η , 使得对所有的 $k, \mathbf{B}_k \leq \eta \mathbf{I}$, 则 $\liminf_k \mathbf{g}_k = 0$.

证明 假设结论不成立, 则存在正常数 η_1 , 使得对所有的 $k, \mathbf{g}_k^T \mathbf{g}_k \geq \eta_1$, 容易证得 $\text{Pre } \mathbf{d}_k \leq \frac{1}{2} \min\{\|\mathbf{g}_k\|, \eta_1\} > 0$, $f_{l(k)} - f_{k+1} \leq \min\{\eta_1 - \eta, \eta_1 \min\{\|\mathbf{g}_k\|, \eta_1\}\}$.

又由水平集 $\mathcal{L}(\mathbf{x}_0)$ 有界, 则 $\{f_k\}$ 有下界, 从而

$$\lim_k f_{l(k)} = \lim_k f_{k+1}. \tag{4}$$

由(3) 和(4) 式可得 $\lim_k \eta_k = 0$. 可以证明对所有的 k ,

$$r_k = \frac{f_{l(k)} - f_{k+1}}{\text{Pre } \mathbf{d}_k} = \frac{f_k - f_{k+1}}{\text{Pre } \mathbf{d}_k},$$
$$\frac{f_k - f_{k+1}}{\text{Pre } \mathbf{d}_k} \rightarrow 1 \quad \frac{o(\|\mathbf{d}_{k(i)}\|^2)}{\frac{1}{2} \mathbf{g}_k^T \min\{\frac{\mathbf{d}_{k(i)}}{\|\mathbf{d}_{k(i)}\|}, \frac{\mathbf{g}_k}{\|\mathbf{g}_k\|}\}}. \tag{5}$$

令 $k \rightarrow \infty$, 有 $\lim_x \frac{f_k - f_{k+1}}{\text{Pre } \mathbf{d}_k} = 1$. 由(5) 式可得, 对充分大的 $k, \frac{f_{l(k)} - f_{k+1}}{\text{Pre } \mathbf{d}_k} \rightarrow 1$. 因此由算法得, 对充分大的 $k, \eta_{k+1} \rightarrow \eta_k$, 与 $\lim_k \eta_k = 0$ 矛盾. 则假设错误, 原定理成立.

定理 3 假设算法产生的点列 $\{\mathbf{x}_k\}$ 收敛到 $\mathbf{x}^*$, $\nabla^2 f(\mathbf{x})$ 在 $\mathbf{x}^*$ 附近连续, 且 $\nabla^2 f(\mathbf{x}^*)$ 正定, 当 $\lim_k \frac{(\nabla^2 f(\mathbf{x}^*) - \mathbf{B}_k) \mathbf{d}_k}{\mathbf{d}_k} = 0$ 时, $\{\mathbf{x}_k\}$ 超线性收敛于 $\mathbf{x}^*$.

证明 由假设条件可得

$$\frac{\mathbf{g}_{k+1}}{\mathbf{d}_k} = \frac{(\mathbf{g}_{k+1} - \mathbf{g}_k) - \frac{1}{2}(\mathbf{g}_k - \mathbf{g}^*)}{\mathbf{d}_k} + \frac{\mathbf{g}_{k+1} - \mathbf{g}_k}{\mathbf{d}_k} + \frac{\frac{1}{2}(\mathbf{g}_k - \mathbf{g}^*)}{\mathbf{d}_k}$$

$$= \frac{(\mathbf{g}_{k+1} - \mathbf{g}_k) - \frac{1}{2}(\mathbf{g}_k - \mathbf{g}^*)}{\mathbf{d}_k} + \frac{1}{2}(\mathbf{g}_k - \mathbf{g}^*) + \frac{\mathbf{g}_{k+1} - \mathbf{g}_k}{\mathbf{d}_k}.$$

其中 $\frac{1}{2}(\mathbf{g}_k - \mathbf{g}^*)$ 为常数.

由 $\lim_{k \rightarrow \infty} \mathbf{x}_k = \mathbf{x}^*$, 则有 $\lim_{k \rightarrow \infty} \frac{\mathbf{g}_{k+1}}{\mathbf{d}_k} = 0$. 又由 $\lim_{k \rightarrow \infty} \mathbf{d}_k = 0$, 则有 $\mathbf{g}(\mathbf{x}^*) = \lim_{k \rightarrow \infty} \mathbf{g}(\mathbf{x}_k) = 0$. 再由 $\frac{1}{2}f(\mathbf{x}^*)$ 非奇异, 则存在 $\delta > 0$, $k_0 \in \mathbb{N}$, 使得对 $k > k_0$, 有 $\mathbf{g}(\mathbf{x}_{k+1}) = \mathbf{g}(\mathbf{x}_{k+1}) - \mathbf{g}(\mathbf{x}^*) = \mathbf{x}_{k+1} - \mathbf{x}^*$. 因此 $\frac{\mathbf{g}_{k+1}}{\mathbf{d}_k} = \frac{\mathbf{g}(\mathbf{x}_{k+1})}{\mathbf{x}_{k+1} - \mathbf{x}_k} = \frac{\mathbf{x}_{k+1} - \mathbf{x}^*}{\mathbf{x}_{k+1} - \mathbf{x}_k + \mathbf{x}_k - \mathbf{x}^*}$, 则有 $\lim_{k \rightarrow \infty} \frac{\mathbf{x}_{k+1} - \mathbf{x}^*}{\mathbf{x}_k - \mathbf{x}^*} = 0$, 即 $\{\mathbf{x}_k\}$ 超线性收敛于 $\mathbf{x}^*$.

参考文献:

- [1] SARTENAER A. Automatic Determination of an Initial Trust Region in Nonlinear Programming [J]. Journal on Scientific computing, 1997, 18(5): 1788–1803.
- [2] ZHANG X S, CHEN Z W, LIAO L Z. A Self-Adaptive Trust Region Method for Unconstrained Optimization [J]. Operate Reserch Transactions, 2001, 5(2): 53–62.
- [3] ZHANG X S, ZHANG J L, LIAO Z. An Adaptive Trust Region Method and Its Convergence [J]. Science in China: Series A, 2002, 45(4): 620–631.
- [4] 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.
- [5] 姚升保. 一类带线搜索的非单调信赖域算法 [J]. 数学杂志, 2003, 23(3): 290–294.
- [6] 李 红, 焦宝聪. 一类带线搜索的非单调自适应信赖域算法 [J]. 首都师范大学学报: 自然科学版, 2008, 29(2): 1–5.
- [7] 李改弟. 一个自动确定信赖域半径的信赖域方法 [J]. 工程数学学报, 2006, 23(5): 843–848.
- [8] 宇振盛, 罗成新. 带记忆信赖域方法的收敛性分析 [J]. 运筹与管理, 2004, 13(2): 16–19.
- [9] FU J H, SUN W. Nonmonotone Adaptive Trust-Region Method for Unconstrained Optimization Problems [J]. Applied Mathematics and Computation, 2005, 163(5): 489–504.
- [10] 陈小军, 张 超. 非线性方程组在几类计算问题中的应用 [J]. 长沙理工大学学报: 自然科学版, 2006, 3(4): 1–7.
- [11] 袁亚湘, 孙文瑜. 最优化理论与方法 [M]. 北京: 科学出版社, 1997.
- [12] 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.
- [13] BACHEM A, GROTSCH M, KORTE B. Mathematical Programming: The State of Art [M]. Berlin: Springer-Verlag, 1983: 258–287.
- [14] 袁亚湘. 信赖域方法的收敛性 [J]. 计算数学, 1994, 16(3): 333–346.
- [15] BULEAU J P, VIAL J P. Curvilinear Path and Trust Region in Unconstrained Optimization, a Convergence Analysis [J]. Math. Prog. Study, 1987, 30: 82–101.
- [16] BYRD B H, SCHNABEL R B, SHULTZ G A. Approximate Solution of the Trust Region Problem by Minimization Over Two-Dimensional Subspace [J]. Math. Prog., 1988, 40: 247–263.

A Nonmonotone Adaptive-BFGS Trust-Region Method

JING Shu-jie, ZHANG Xiao-liang

(Mathematics and Information Science, He nan Polytechnic University, Jiaozuo 454003, He nan China)

Abstract: A nonmonotone adaptive-BFGS trust-region method for unconstrained optimization is presented. Not only the trust-region radius in this method is automatically determined by using first order information, but combining with the advantage of the BFGS algorithm. Under certain conditions, the global and superlinear convergences of the algorithm are proved.

Key words: unconstrained optimization; nonmonotone algorithm; adaptive trust algorithm; BFGS update; global convergence

(责任编辑 向阳洁)