

一个锥模型的自适应信赖域算法及其收敛性

冯 琳¹, 段复建²

(1. 重庆文理学院数学与财经学院, 重庆 402160)

(2. 桂林电子科技大学数学与计算科学学院, 广西 桂林 541004)

摘要: 本文研究了无约束最优化问题的基于锥模型的自适应信赖域算法. 利用理论分析得到一个新的自适应信赖域半径. 算法在每步迭代中以变化的速率、当前迭代点的信息以及水平向量信息调节信赖域半径的大小. 从理论上证明了新算法的全局收敛性和 Q -二阶收敛性. 用数值试验验证了新算法的有效性. 推广了已有的自适应信赖域算法的可行性和有效性.

关键词: 无约束最优化; 信赖域方法; 锥模型; 自适应; 收敛性

MR(2010) 主题分类号: 90C30; 65K10

中图分类号: 0224.2

文献标识码: A

文章编号: 0255-7797(2016)01-0144-13

1 引言

考虑无约束最优化问题

$$\min_{x \in R^n} f(x), \quad (1.1)$$

其中 $f(x) : R^n \rightarrow R^1$ 是连续可微函数. 本文采用如下记号:

· $\|\cdot\|$ 表示 Euclidean 范数.

· $g(x) \in R^n$ 是 $f(x)$ 在点 x 的梯度.

· $H(x) \in R^{n \times n}$ 是 $f(x)$ 在点 x 的 Hessian 矩阵.

· $\{x_k\}$ 是某一算法产生的迭代点列, 并记 $f_k = f(x_k)$, $g_k = g(x_k)$, $H_k = H(x_k)$.

· $B_k \in R^{n \times n}$ 是对称矩阵, 它是 $f(x)$ 在点 x_k 的 Hessian 矩阵或其近似.

· \bar{B}_k 是安全正定矩阵且满足 Schnabel 和 Eskow^[1] 改进的 Cholesky 分解. $\bar{B}_k = B_k + E_k$, E_k 是一个使得 \bar{B}_k 安全正定的对角矩阵; 当 B_k 正定时, 取 $E_k = 0$.

1970 年, Powell^[2] 首次提出求解问题 (1.1) 的信赖域算法, 其思想是通过求解一系列二次函数在信赖域中的最优点的方法来求问题 (1.1) 的解. 信赖域算法的显著特点是其稳定的数值性能、强收敛性和强适性, 能有效地解决病态问题, 需要迭代的次数少, 因此信赖域算法在非线优化界受到了特别的重视, 特别是最近几年一直是非线性优化界研究的一个热点. 目前信赖域算法已经和线搜索方法并列为求解非线性优化问题的两类主要的数值方法.

信赖域算法是在每一步迭代, 求解信赖域子问题

$$\min q_k(s) = f_k + g_k^T s + \frac{1}{2} s^T B_k s \quad \text{s.t. } \|s\| \leq \Delta_k, \quad (1.2)$$

*收稿日期: 2011-10-13 接收日期: 2012-03-01

基金项目: 国家自然科学基金 (11061011); 广西自然科学基金 (2011GXNSFA018138); 重庆文理学院课题 (Y2013sc42).

作者简介: 冯琳 (1973-), 女, 河南鹤壁, 讲师, 主要研究方向: 最优化理论与方法.

其中 $s = x - x_k$, $g_k = \nabla f(x_k)$, $\Delta_k > 0$ 是信赖域半径.

在当前迭代点 x_k , 设子问题 (1.2) 的解是 s_k . $f(x)$ 在点 x_k 处的预估下降量是

$$\text{Pres}_k = -g_k^T s_k - \frac{1}{2} s_k^T B_k s_k;$$

实际下降量是

$$\text{Ares}_k = f(x_k) - f(x_k + s_k).$$

实际下降量 Ares_k 与预估下降量 Pres_k 的比值

$$r_k = \frac{\text{Ares}_k}{\text{Pres}_k},$$

它反映了子问题 (1.2) 的解 s_k 令人满意的程度, r_k 越接近于 1, 表明 s_k 越令人满意. 如果 s_k 令人满意, 就接受 s_k , 得到新的迭代点 x_{k+1} , 同时增大信赖域半径 Δ_k ; 反之, 则拒绝 s_k , 并需缩小信赖域半径 Δ_k , 重解子问题 (1.2), 直至 s_k 被接受. 即

$$x_{k+1} = \begin{cases} x_k + s_k, & r_k \geq \mu, \\ x_k, & r_k < \mu, \end{cases}$$

其中 $\mu \in (0, 1)$ 是一个常数, 并修正信赖域半径

$$\Delta_{k+1} = \begin{cases} c_1 \Delta_k, & r_k < \mu_1, \\ c_2 \Delta_k, & r_k > \mu_2, \\ \Delta_k, & \mu_1 \leq r_k \leq \mu_2, \end{cases} \quad (1.3)$$

其中 $0 \leq \mu_1 < \mu_2 < 1, 0 < c_1 < 1 < c_2$ 是常数.

(1.3) 式对 Δ_k 的修正没有利用对算法有重大影响的 g_k, B_k 等这些当前迭代点的信息, 而只是根据 r_k 按常数倍放大或缩小初始信赖域半径, 这降低了算法的效率. 基于此, 许多自适应信赖域方法被提出. Zhang et al. [3] 给出一个自适应信赖域方法, 信赖域半径取为 $\Delta_k = c^p \|g_k\| \overline{M}_k$ (这里 $0 < c < 1, p$ 是非负整数变量, $\overline{M}_k = \|B_k^{-1}\|$), 如果上次迭代的试探步 s_{k-1} 被接受, 则 $p = 0$; 否则, $p = p + 1$. 姚升保等 [4] 提出信赖域半径取为 $\Delta_{k+1} \in [r_1 \|d_k\|, r_2 \|d_k\|]$ 或 $\Delta_{k+1} \in [\|d_k\|, r_3 \|d_k\|]$ (这里 $0 < r_1 < r_2 < 1 < r_3$) 的自适应信赖域方法. 文 [5] 也给出了一个自适应信赖域算法, 信赖域半径取为 $\Delta_k = R_\eta(t) \|d_{k-1}\|$, 其中 $R_\eta(t)$ 是一个关于 r_k 的 R -函数.

定义 1.1 [5] $R_\eta(t)$ 定义在 $(-\infty, +\infty)$ 上, 参数 $\eta \in (0, 1)$, $R_\eta(t)$ 是一个 R -函数当且仅当满足:

- (i) $R_\eta(t)$ 在 $(-\infty, +\infty)$ 非减;
- (ii) $\lim_{t \rightarrow -\infty} R_\eta(t) = \beta$ ($\beta \in (0, 1)$ 是一个常数);
- (iii) $R_\eta(t) \leq 1 - \gamma_1$ ($\forall t < \eta, \gamma_1 \in (0, 1 - \beta)$ 是一个常数);
- (iv) $R_\eta(\eta) = 1 + \gamma_2$ ($\gamma_2 \in (0, \beta)$ 是一个常数);
- (v) $\lim_{t \rightarrow +\infty} R_\eta(t) = M$ ($M \in (1 + \gamma_2, +\infty)$ 是一个常数).

由定义 1.1, 我们得到 R -函数的一些性质:

定理 1.1 [5] 如果 $R_\eta(t)$ ($\eta \in (0, 1)$) 是一个 R -函数, 则

$$0 < \beta \leq R_\eta(t) \leq 1 - \gamma_1 < 1, \forall t \in (-\infty, \eta), \quad (1.4)$$

$$1 < 1 + \gamma_2 \leq R_\eta(t) \leq M < +\infty, \forall t \in [\eta, +\infty), \quad (1.5)$$

因此可以把 $R_\mu(r_k)$ 作为增大或缩小信赖域半径的尺度, 使信赖域半径的调节依据于问题本身, 更有客观性.

大多数信赖域算法采用二次模型逼近 $f(x)$, 但对一些非二次性态强, 曲率变化剧烈的函数, 用二次模型逼近 $f(x)$ 效果较差, 因而得到的最优点较差. 1980 年, Davidon [6] 首次提出锥模型. 锥模型比二次模型更一般, 包含的信息和自由度更多, 能更好地逼近 $f(x)$, 因此许多学者对它进行了研究. Sorensen [7], Ariyawansa [8] 和 Sheng [9] 研究了锥模型共线调比拟牛顿方法. 鉴于锥模型和信赖域算法各具优点, 许多学者将二者结合进行研究. 1995 年, 诸梅芳, 薛毅等 [10] 提出求解问题 (1.1) 的锥模型信赖域算法; 2005 年, Ni 等 [11] 给出锥模型信赖域子问题的最优性条件, 为锥模型信赖域子问题的求解提供了更充分且合理的理论基础. 信赖域算法进一步发展. 2005 年, Fu 等 [12] 将自适应技术引入锥模型信赖域算法, 提出锥模型自适应信赖域算法; 2010 年, 王希云等 [13] 也提出锥模型自适应信赖域算法. 锥模型信赖域算法得到了广泛的发展, 并且数值结果表明对一些函数, 特别是对曲率变化剧烈的函数, 锥模型信赖域算法比二次模型信赖域算法效果更好.

求解问题 (1.1) 的一个典型的锥模型信赖域子问题是

$$\min_{s_k \in R^n} q_k(s_k) = f_k + \frac{g_k^T s_k}{1 + b_k^T s_k} + \frac{1}{2} \frac{s_k^T B_k s_k}{(1 + b_k^T s_k)^2} \quad \text{s.t.} \quad \|s_k\| \leq \Delta_k, \quad (1.6)$$

其中 $b_k \in R^n$ 是水平向量, $1 + b_k^T s_k > 0$. 当 $b_k = 0$ 或 $b_k^T s_k = 0$ 时, 锥模型转化为二次模型, 因此锥模型是二次模型的推广.

鉴于文 [3] 和文 [5] 工作的基础上, 基于锥模型信赖域子问题 (1.6), 本文提出了一类新的锥模型自适应信赖域算法, 信赖域半径取为

$$\Delta_k = R_\mu(r_k) \frac{\|B_k^{-1}\| \|g_k\|}{|1 + b_k^T B_k^{-1} g_k|}, \quad (1.7)$$

使得在每步迭代中, 信赖域半径的修正依据问题本身确定.

本文其他部分安排如下: 第 2 节给出了锥模型自适应信赖域算法, 第 3, 4 节分别分析了算法的全局收敛性和 Q -二阶收敛速度, 最后一节给出了数值实验.

2 锥模型自适应信赖域算法

在本节中, 基于锥模型信赖域子问题 (1.6) 和信赖域半径 (1.7), 我们提出一个求解问题 (1.1) 的自适应信赖域算法.

设问题 (1.6) 的解为 s_k , 目标函数 $f(x)$ 在第 k 步的实际下降量为

$$\text{Ares}_k = f_k - f(x_k + s_k), \quad (2.1)$$

预估下降量为

$$\text{Pres}_k = q_k(0) - q_k(s_k) = -\frac{g_k^T s_k}{1 + b_k^T s_k} - \frac{1}{2} \frac{s_k^T B_k s_k}{(1 + b_k^T s_k)^2}, \quad (2.2)$$

比值

$$r_k = \frac{\text{Ares}_k}{\text{Pres}_k}. \quad (2.3)$$

下面给出新的锥模型自适应信赖域算法 (CSATR) 的具体实现过程:

步 0 给出初始点 $x_0 \in R^n$, $\Delta_0 > 0$, $b_0 \in R^{n \times 1}$, $B_0 = I$ (单位阵), $0 < \beta < 1$, $0 < \gamma_1 < 1 - \beta$, $\gamma_2 > 0$, $M > 1 + \gamma_2$, $\varepsilon > 0$, $0 < \mu < 1$, $k := 0$.

步 1 计算 $g_k = g(x_k)$. 如果 $\|g_k\| \leq \varepsilon$, 则停止计算, $x^* = x_k$; 否则, 转步 2.

步 2 近似求解问题 (1.6) 得到 s_k .

步 3 利用 (2.1), (2.2), (2.3) 式分别计算 Ares_k , Pres_k , r_k .

步 4 如果 $r_k < \mu$, 令 $x_{k+1} = x_k$; 否则, 令 $x_{k+1} = x_k + s_k$.

步 5 修正 b_k 及 B_k , 产生 b_{k+1} 及 B_{k+1} , 利用 (1.7) 式计算 Δ_{k+1} . 令 $k := k + 1$, 转步 1.

注 (i) 算法 (CSATR) 中, 要求 $\|b_k\|\Delta_k < 1$; 如果 $\|b_k\|\Delta_k \geq 1$, 则令 $\Delta_k = \frac{\alpha}{\|b_k\|}$ ($0 < \alpha < 1$) 使得 $\|b_k\|\Delta_k < 1$.

(ii) 算法 (CSATR) 步 2 中 s_k 的具体求解及步 5 中 b_k 和 B_k 的修正见文 [12].

(iii) 若 (1.6) 及 (1.7) 式中的 $b_k = 0$, 则算法 (CSATR) 转化为相应的二次模型自适应信赖域算法 (QSATR).

(iv) 算法 (CSATR) 中, 利用 (1.3) 式更新信赖域半径, 得到传统的锥模型信赖域算法 (CTNTR).

3 算法的全局收敛性

本文所需假设如下:

A1: $f(x)$ 在水平集 $L(x_0) = \{x | f(x) \leq f(x_0)\}$ 上有下界.

A2: $\{B_k\}$, $\{B_k^{-1}\}$, $\{b_k\}$ 一致有界, 即存在常数 $\delta_1, \delta_2, \delta_3 > 0$, 使得 $\forall k$, 有 $\|B_k\| \leq \delta_1$, $\|B_k^{-1}\| \leq \delta_2$, $\|b_k\| \leq \delta_3$.

A3: $g(x)$ 在 $L(x_0)$ 上一致连续.

本文算法 (CSATR) 中的 $B_k, \forall k$, 用文 [12] 中的方法修正时保持正定性. 以下的 $B_k, \forall k$ 均为对称正定矩阵.

为讨论方便, 记 $I = \{k : r_k \geq \mu\}$, $J = \{k : r_k < \mu\}$.

引理 3.1 设 s_k 是子问题 (1.6) 的解, $g_k \neq 0$, 则

$$q_k(0) - q_k(s_k) > 0. \quad (3.1)$$

为得到算法的下降性条件, 引入 Cauchy 点

$$S_k^C = -T_k \frac{\Delta_k}{\|g_k\|} g_k,$$

其中

$$T_k = \begin{cases} \min\{1, \frac{\|g_k\|^3}{\Delta_k (g_k^T B_k g_k + g_k^T g_k b_k^T g_k)}\}, & g_k^T B_k g_k + g_k^T g_k b_k^T g_k > 0, \\ 1, & g_k^T B_k g_k + g_k^T g_k b_k^T g_k \leq 0. \end{cases}$$

引理 3.2 对 Cauchy 点 S_k^C 满足

$$q_k(0) - q_k(S_k^C) \geq \frac{1}{2} \|g_k\| \min\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\}. \quad (3.2)$$

证 分三种情况证明.

(i) 当 $g_k^T B_k g_k + g_k^T g_k b_k^T g_k > 0$ 且 $1 \geq \frac{\|g_k\|^3}{\Delta_k (g_k^T B_k g_k + g_k^T g_k b_k^T g_k)}$ 时, $T_k = \frac{\|g_k\|^3}{\Delta_k (g_k^T B_k g_k + g_k^T g_k b_k^T g_k)}$, $S_k^C = -\frac{\|g_k\|^2}{g_k^T B_k g_k + g_k^T g_k b_k^T g_k}$, 则

$$\begin{aligned} q_k(0) - q_k(S_k^C) &= -\frac{g_k^T S_k^C}{1 + b_k^T S_k^C} - \frac{1}{2} \frac{S_k^{CT} B_k S_k^C}{(1 + b_k^T S_k^C)^2} = \frac{\|g_k\|^4}{g_k^T B_k g_k} - \frac{1}{2} \frac{\|g_k\|^4}{g_k^T B_k g_k} \\ &= \frac{1}{2} \frac{\|g_k\|^4}{g_k^T B_k g_k} \geq \frac{1}{2} \frac{\|g_k\|^2}{\|B_k\|} \geq \frac{1}{2} \|g_k\| \min\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\}. \end{aligned}$$

(ii) 当 $g_k^T B_k g_k + g_k^T g_k b_k^T g_k > 0$ 且 $1 < \frac{\|g_k\|^3}{\Delta_k (g_k^T B_k g_k + g_k^T g_k b_k^T g_k)}$ 时,

$$\|g_k\|^3 - \Delta_k g_k^T g_k b_k^T g_k - \Delta_k g_k^T B_k g_k > 0, \quad (3.3)$$

$$g_k^T B_k g_k + g_k^T g_k b_k^T g_k > 0, \quad (3.4)$$

$$S_k^C = -\frac{\Delta_k}{\|g_k\|} g_k. \quad (3.5)$$

由 $\|b_k\| \Delta_k < 1$ 知 $\Delta_k b_k^T g_k \leq \Delta_k \|b_k\| \|g_k\| < \|g_k\|$, 则

$$\|g_k\| - \Delta_k b_k^T g_k > 0. \quad (3.6)$$

由 (3.3)–(3.6) 式知

$$\begin{aligned} q_k(0) - q_k(S_k^C) &= -\frac{g_k^T S_k^C}{1 + b_k^T S_k^C} - \frac{1}{2} \frac{S_k^{CT} B_k S_k^C}{(1 + b_k^T S_k^C)^2} = \frac{\Delta_k g_k^T g_k}{\|g_k\| - \Delta_k b_k^T g_k} - \frac{\Delta_k^2 g_k^T B_k g_k}{2(\|g_k\| - \Delta_k b_k^T g_k)^2} \\ &= \frac{\Delta_k [\|g_k\|^3 - \Delta_k g_k^T g_k b_k^T g_k - \Delta_k g_k^T B_k g_k] + \Delta_k g_k^T g_k (\|g_k\| - \Delta_k b_k^T g_k)}{2(\|g_k\| - \Delta_k b_k^T g_k)^2} \\ &\geq \frac{\Delta_k g_k^T g_k}{2(\|g_k\| - \Delta_k b_k^T g_k)} \geq \frac{\Delta_k g_k^T g_k}{2(\|g_k\| + \Delta_k \|b_k^T\| \|g_k\|)} \geq \frac{1}{2} \frac{\Delta_k \|g_k\|^2}{\|g_k\|} \\ &= \frac{1}{2} \Delta_k \|g_k\| \geq \frac{1}{2} \|g_k\| \min\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\}. \end{aligned}$$

(iii) 当 $g_k^T B_k g_k + g_k^T g_k b_k^T g_k \leq 0$ 时,

$$g_k^T B_k g_k + g_k^T g_k b_k^T g_k \leq 0, \quad (3.7)$$

$$S_k^C = -\frac{\Delta_k}{\|g_k\|} g_k. \quad (3.8)$$

由 (3.7) 式知 $0 < g_k^T B_k g_k \leq -g_k^T g_k b_k^T g_k = \|g_k\|^2 b_k^T g_k$. 于是

$$b_k^T g_k < 0. \quad (3.9)$$

从而

$$\|g_k\| - \Delta_k b_k^T g_k > 0. \quad (3.10)$$

由 (3.7) 式知

$$g_k^T B_k g_k + g_k^T g_k b_k^T g_k \leq \frac{g_k^T g_k \|g_k\|}{\Delta_k}.$$

于是

$$g_k^T B_k g_k \leq \frac{g_k^T g_k \|g_k\| - \Delta_k g_k^T g_k b_k^T g_k}{\Delta_k}. \quad (3.11)$$

又

$$\frac{g_k^T g_k \|g_k\| - \Delta_k g_k^T g_k b_k^T g_k}{\Delta_k} = \frac{g_k^T g_k (\|g_k\| - \Delta_k b_k^T g_k)}{\Delta_k}, \quad (3.12)$$

由 (3.11) 和 (3.12) 式知

$$0 < \frac{\Delta_k g_k^T B_k g_k}{(\|g_k\| - \Delta_k b_k^T g_k)^2} \leq \frac{g_k^T g_k}{\|g_k\| - \Delta_k b_k^T g_k}. \quad (3.13)$$

于是由 (3.8), (3.13) 式, (ii) 的过程, (3.9) 和 (3.10) 式知

$$\begin{aligned} q_k(0) - q_k(S_k^C) &= \frac{\Delta_k g_k^T g_k}{\|g_k\| - \Delta_k b_k^T g_k} - \frac{\Delta_k^2 g_k^T B_k g_k}{2(\|g_k\| - \Delta_k b_k^T g_k)^2} \\ &\geq \frac{\Delta_k g_k^T g_k}{\|g_k\| - \Delta_k b_k^T g_k} - \frac{\Delta_k g_k^T g_k}{2\|g_k\| - \Delta_k b_k^T g_k} \\ &= \frac{1}{2} \frac{\Delta_k g_k^T g_k}{\|g_k\| - \Delta_k b_k^T g_k} \geq \frac{1}{2} \frac{\Delta_k \|g_k\|^2}{\|g_k\|} \\ &= \frac{1}{2} \Delta_k \|g_k\| \geq \frac{1}{2} \|g_k\| \min\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\}. \end{aligned}$$

证毕.

下面的引理说明子问题 (1.6) 的解 s_k 满足充分下降条件.

引理 3.3 设 s_k 是子问题 (1.6) 的解, 则

$$\text{Pres}_k \geq \frac{1}{2} \|g_k\| \min\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\}. \quad (3.14)$$

证 由 (3.2) 式知

$$\text{Pres}_k = q_k(0) - q_k(s_k) \geq q_k(0) - q_k(S_k^C) \geq \frac{1}{2} \|g_k\| \min\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\}.$$

引理 3.4 如果假设 A1–A3 成立, 则存在常数 $c_0 > 0$, 使得

$$\|s_k\| \leq c_0 \|g_k\|. \quad (3.15)$$

证 由假设 A3 知存在 $0 < \varepsilon < \lambda$, 使得 $\varepsilon \leq \|g_k\| \leq \lambda$. 于是由假设 A2 知

$$|1 + b_k^T B_k^{-1} g_k| \geq 1 - |b_k^T B_k^{-1} g_k| \geq 1 - \|b_k^T\| \|B_k^{-1}\| \|g_k\| \geq 1 - \delta_3 \delta_2 \lambda > 0. \quad (3.16)$$

于是由 (1.5) 式, 假设 A2 和 (3.16) 式知

$$\Delta_k = R_\mu(r_k) \frac{\|B_k^{-1}\| \|g_k\|}{|1 + b_k^T B_k^{-1} g_k|} \leq \frac{M \|B_k^{-1}\| \|g_k\|}{|1 + b_k^T B_k^{-1} g_k|} \leq \frac{M \delta_2}{1 - \delta_3 \delta_2 \lambda} \|g_k\|.$$

令 $c_0 = \frac{M \delta_2}{1 - \delta_3 \delta_2 \lambda}$, 则 $\Delta_k \leq c_0 \|g_k\|$. 从而 $\|s_k\| \leq c_0 \|g_k\|$.

引理 3.5 如果 $\{x_k\}$ 是由算法 (CSATR) 产生的点列, 则 $\{x_k\} \subset L(x_0)$.

证 用数学归纳法证明.

当 $k = 0$ 时, $f(x_0) \leq f(x_0)$, 则 $x_k \in L(x_0)$, 命题成立.

下证假设 $x_k \in L(x_0)$ 时, 则有 $x_{k+1} \in L(x_0)$. 当 $x_k \in L(x_0)$ 时, 有 $f(x_k) \leq f(x_0)$.

(i) 如果 $k \in I$, 则 $r_k \geq \mu$.

于是由 (3.1) 式知 $f(x_k) - f(x_{k+1}) \geq \mu(q_k(0) - q_k(s_k)) > 0$. 因而 $f(x_{k+1}) \leq f(x_k)$. 所以 $f(x_{k+1}) \leq f(x_0)$.

(ii) 如果 $k \in J$, 则 $r_k < \mu$.

由算法 (CSATR) 步 4 知 $x_{k+1} = x_k$, 于是 $f(x_{k+1}) = f(x_k)$. 所以 $f(x_{k+1}) \leq f(x_0)$.

由 (i) 和 (ii) 知当 $x_k \in L(x_0)$ 时, $x_{k+1} \in L(x_0)$.

由以上知结论成立.

算法 (CSATR) 中, 如果 $x_{k+1} = x_k + s_k$, 则 x_{k+1} 是一个成功的迭代点; 如果 $x_{k+1} = x_k$, 则 x_{k+1} 是一个非成功的迭代点.

引理 3.6 假设 A1, A3 成立, $\{x_k\}$ 是由算法 (CSATR) 产生的无穷点列, 且 $\forall k$, 有 $\|g_k\| > \varepsilon$, $\varepsilon \in (0, 1)$ 是常数, 则对 $\forall k$, 存在非负整数 p 使得 x_{k+p+1} 是一个成功的迭代点.

证 假设存在 k_0 使得 $\forall p$ 都有 x_{k_0+p+1} 是一个非成功的迭代点. 即

$$r_{k_0+p} < \mu, p = 0, 1, 2, \dots \quad (3.17)$$

由算法 (CSATR) 步 4 知

$$x_{k_0+p+1} = x_k, p = 0, 1, 2, \dots \quad (3.18)$$

由 (1.7) 和 (1.4) 式知

$$\Delta_{k_0+p+1} \rightarrow 0, p \rightarrow \infty, \quad (3.19)$$

于是

$$\|s_{k_0+p+1}\| \rightarrow 0, p \rightarrow \infty.$$

因而对充分大的 p , 有

$$\frac{1}{1 + b_{k_0+p}^T s_{k_0+p}} = 1 + O(\|s_{k_0+p}\|),$$

于是

$$\frac{g_{k_0+p}^T s_{k_0+p}}{1 + b_{k_0+p}^T s_{k_0+p}} = g_{k_0+p}^T s_{k_0+p} + O(\|s_{k_0+p}\|^2), \quad (3.20)$$

$$\frac{s_{k_0+p}^T B_{k_0+p} s_{k_0+p}}{(1 + b_{k_0+p}^T s_{k_0+p})^2} = s_{k_0+p}^T B_{k_0+p} s_{k_0+p} + o(\|s_{k_0+p}\|^2). \quad (3.21)$$

由 (3.20) 和 (3.21) 式知

$$\begin{aligned} q_{k_0+p}(0) - q_{k_0+p}(s_{k_0+p}) &= -\frac{g_{k_0+p}^T s_{k_0+p}}{1 + b_{k_0+p}^T s_{k_0+p}} - \frac{1}{2} \frac{s_{k_0+p}^T B_{k_0+p} s_{k_0+p}}{(1 + b_{k_0+p}^T s_{k_0+p})^2} \\ &= -g_{k_0+p}^T s_{k_0+p} - \frac{1}{2} s_{k_0+p}^T B_{k_0+p} s_{k_0+p} + o(\|s_{k_0+p}\|^2). \end{aligned} \quad (3.22)$$

由 (3.14) 和 (3.19) 式知

$$\begin{aligned} q_{k_0+p}(0) - q_{k_0+p}(s_{k_0+p}) &\geq \frac{1}{2} \|g_{k_0+p}\| \min\{\Delta_{k_0+p}, \frac{\|g_{k_0+p}\|}{\|B_{k_0+p}\|}\} \geq \frac{1}{2} \|g_{k_0+p}\| \Delta_{k_0+p} \\ &\geq \frac{1}{2} \|g_{k_0+p}\| \|s_{k_0+p}\| \geq \frac{1}{2} \varepsilon \|s_{k_0+p}\|. \end{aligned} \quad (3.23)$$

由 $f(x)$ 是连续可微函数及 Taylor 展开式知

$$f_{k_0+p} - f(x_{k_0+p} + s_{k_0+p}) = -g_{k_0+p}^T s_{k_0+p} - \frac{1}{2} s_{k_0+p}^T B_{k_0+p} s_{k_0+p} + o(\|s_{k_0+p}\|^2). \quad (3.24)$$

由 (3.22)–(3.24) 式知

$$\left| \frac{f_{k_0+p} - f(x_{k_0+p} + s_{k_0+p})}{q_{k_0+p}(0) - q_{k_0+p}(s_{k_0+p})} - 1 \right| \leq \frac{o(\|s_{k_0+p}\|^2)}{\frac{1}{2} \varepsilon \|s_{k_0+p}\|} \rightarrow 0. \quad (3.25)$$

因此当 p 充分大时, 由 (3.25) 式及 $0 < \mu < 1$ 知 $r_{k_0+p} \geq \mu$, 这与 (3.17) 式矛盾.

定理 3.1 如果假设 A1 成立, $\{x_k\}$ 是由算法 (CSATR) 产生的点列, 则

$$\lim_{k \rightarrow \infty} \|g_k\| = 0. \quad (3.26)$$

证 (反证法) 假设结论不成立, 则存在常数 $\varepsilon > 0, \varepsilon \in (0, 1)$, 使得 $\|g_k\| \geq \varepsilon, \forall k$.

下证

$$\lim_{k \in I, k \rightarrow \infty} R_\mu(r_k) = 0. \quad (3.27)$$

由引理 3.6 知当 k 充分大时, 则 $k \in I$, 此时有

$$f_k - f(x_k + s_k) > \mu(q_k(0) - q_k(s_k)),$$

由于 $\{f_k\}$ 单调递减有下界, 因此 $\{f_k\}$ 收敛.

于是当 k 充分大时, 有

$$q_k(0) - q_k(s_k) \rightarrow 0. \quad (3.28)$$

由 (3.14), (1.7), (3.16) 式和假设 A2 知

$$\begin{aligned} q_k(0) - q_k(s_k) &\geq \frac{1}{2} \|g_k\| \min\{\Delta_k, \frac{\|g_k\|}{\|B_k}\|}\} \geq \frac{1}{2} \varepsilon \min\{R_\mu(r_k) \frac{\|B_k^{-1}\| \|g_k\|}{|1 + b_k^T B_k^{-1} g_k|}, \frac{\varepsilon}{\delta_1}\} \\ &\geq \frac{1}{2} \varepsilon \min\{R_\mu(r_k) \frac{\|g_k\|}{|1 + b_k^T B_k^{-1} g_k| \|B_k\|}, \frac{\varepsilon}{\delta_1}\} \\ &\geq \frac{1}{2} \varepsilon \min\{R_\mu(r_k) \frac{\varepsilon}{(1 - \delta_3 \delta_2 \lambda) \delta_1}, \frac{\varepsilon}{\delta_1}\}. \end{aligned} \quad (3.29)$$

由 (3.28) 和 (3.29) 式知

$$\lim_{k \rightarrow \infty} R_\mu(r_k) = 0.$$

又 $k \in I$ 时, 有 $r_k \geq \mu$, 由定义 1.1 和定理 1.1 知 $R_\mu(r_k) \geq 1 + \gamma_2 > 1, \forall k \in I$, 这与 (3.27) 式矛盾. 因此定理成立.

4 超线性收敛性

首先给出三个假设:

A4: 算法 (CSATR) 产生的点列 $\{x_k\}$ 满足 $\lim_{k \rightarrow \infty} x_k = x^*$, $\lim_{k \rightarrow \infty} \|g_k\| = \|g^*\| = 0$, 且 $\nabla^2 f(x^*)$ 正定.

A5: 如果 $\frac{\|B_k^{-1}g_k\|}{|1+b_k^T B_k^{-1}g_k|} \leq \Delta_k$, 则 $s_k = -\frac{B_k^{-1}g_k}{1+b_k^T B_k^{-1}g_k}$.

A6: 算法 (CSATR) 产生的点列 $\{x_k\}$ 及 $\{B_k\}$ 满足

$$\lim_{k \rightarrow \infty} \frac{\|(B_k - \nabla^2 f(x_k)S_s)\|}{\|S_s\|} = 0.$$

定理 4.1 设 $\{x_k\}$ 是由算法 (CSATR) 产生的点列且满足假设 A1–A5, 则点列 $\{x_k\}$ 超线性收敛于 x^* .

证 由假设 A4 知 $\exists W > w > 0$ 使得 $\forall x \in D = \{x \mid \|x - x^*\| \leq \Delta\}$ (其中 Δ 为一个很小的常数) 时, 有

$$w\|y\|^2 \leq y^T \nabla^2 f(x)y \leq W\|y\|^2, \forall y \in R^n$$

且存在充分大的正整数 k_0 , 使得 $\forall k \geq k_0$ 时, 有 $x_k \in D$.

由 Taylor 展开式和假设 A4 知, 当 $\forall k \geq k_0$ 时, 有

$$\frac{1}{2}w\|x_k - x^*\|^2 \leq f_k - f(x^*) \leq \frac{1}{2}W\|x_k - x^*\|^2, \quad (4.1)$$

且

$$w\|x_k - x^*\| \leq \|g_k\| \leq W\|x_k - x^*\|. \quad (4.2)$$

下证对充分大的 $k \geq k_0$, 存在常数 c_1 使得

$$q_k(0) - q_k(s_k) \geq c_1\|s_k\|^2. \quad (4.3)$$

由 (3.14), (3.15) 式和假设 A2 知对充分大的 k , 有

$$\begin{aligned} q_k(0) - q_k(s_k) &\geq \frac{1}{2}\|g_k\| \min\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\} \geq \frac{1}{2} \frac{\|s_k\|}{c_0} \min\{\|s_k\|, \frac{\|s_k\|}{c_0\delta_1}\} \\ &\geq \frac{1}{2c_0}\|s_k\|^2 \min\{1, \frac{1}{c_0\delta_1}\}. \end{aligned}$$

取 $c_1 = \frac{1}{2c_0} \min\{1, \frac{1}{c_0\delta_1}\}$, 于是 (4.2) 式成立.

下证当 k 充分大时, 有 $r_k \geq \mu$.

由 Taylor 展开式和假设 A6 知

$$\begin{aligned}
 & f(x_k) - f(x_k + s_k) - q_k(0) - q_k(s_k) \\
 = & f(x_k) - f(x_k + s_k) + \frac{g_k^T s_k}{1 + b_k^T s_k} + \frac{1}{2} \frac{s_k^T B_k s_k}{(1 + b_k^T s_k)^2} \\
 = & -g_k^T s_k - \frac{1}{2} s_k^T \nabla^2 f(x_k) s_k + o(\|s_k\|)^2 + g_k^T s_k [1 + O(\|s_k\|^2)] + \frac{1}{2} s_k^T B_k s_k [1 + o(\|s_k\|^2)] \\
 = & \frac{1}{2} s_k^T [B_k - \nabla^2 f(x_k)] s_k + o(\|s_k\|^2). \tag{4.4}
 \end{aligned}$$

由 (4.3) 和 (4.4) 式知

$$\left| \frac{f(x_k) - f(x_k + s_k)}{q_k(0) - q_k(s_k)} - 1 \right| = \left| \frac{f(x_k) - f(x_k + s_k) - (q_k(0) - q_k(s_k))}{q_k(0) - q_k(s_k)} - 1 \right| \leq \frac{o(\|s_k\|^2)}{c_1 \|s_k\|^2},$$

因此

$$\frac{f(x_k) - f(x_k + s_k)}{q_k(0) - q_k(s_k)} \rightarrow 1, k \rightarrow \infty.$$

于是当 k 充分大时, 有 $r_k \geq \mu$. 当 $r_k \geq \mu$ 时, 由 (1.5) 式知 $R_\mu(r_k) > 1$. 于是

$$\Delta_k = R_\mu(r_k) \frac{\|B_k^{-1}\| \|g_k\|}{|1 + b_k^T B_k^{-1} g_k|} > \frac{\|B_k^{-1}\| \|g_k\|}{|1 + b_k^T B_k^{-1} g_k|} \geq \frac{\|B_k^{-1} g_k\|}{|1 + b_k^T B_k^{-1} g_k|},$$

则由假设 A5 知

$$s_k = -\frac{B_k^{-1} g_k}{1 + b_k^T B_k^{-1} g_k},$$

于是有

$$g_k + B_k s_k - b_k^T B_k^{-1} g_k B_k s_k = 0. \tag{4.5}$$

由算法 (CSATR) 步 4 及引理 3.6 知当 k 充分大时, 有

$$s_k = x_{k+1} - x_k. \tag{4.6}$$

由 (4.6), (4.1), (4.2) 式及 $\{f_k\}$ 单调递减知

$$\begin{aligned}
 \|s_k\| & \leq \|x_k - x^*\| + \|x_{k+1} - x^*\| \leq \|x_k - x^*\| + \sqrt{\frac{2(f_{k+1} - f(x^*))}{w}} \\
 & \leq \|x_k - x^*\| + \sqrt{\frac{2(f_k - f(x^*))}{w}} \leq (1 + \sqrt{\frac{W}{w}}) \frac{\|g(x_k)\|}{w}. \tag{4.7}
 \end{aligned}$$

由 Taylor 展开式知

$$g_{k+1} = g_k + \nabla^2 f(x_k) s_k + O(\|s_k\|^2). \tag{4.8}$$

由 (4.5) 和 (4.8) 式知

$$\begin{aligned}
 \frac{\|g_{k+1}\|}{\|s_k\|} &= \frac{\|g_{k+1} - g_k - B_k s_k + b_k^T B_k^{-1} g_k B_k s_k\|}{\|s_k\|} \\
 &\leq \frac{\|g_{k+1} - g_k - B_k s_k\| + \|b_k^T\| \|B_k^{-1}\| \|g_k\| \|B_k\| \|s_k\|}{\|s_k\|} \\
 &= \frac{\|g_{k+1} - g_k - B_k s_k\|}{\|s_k\|} + \|b_k^T\| \|B_k^{-1}\| \|g_k\| \|B_k\| \\
 &\leq \frac{\|\nabla^2 f(x_k) - B_k\| \|s_k\|}{\|s_k\|} + \frac{O(\|s_k\|^2)}{\|s_k\|} + \delta_3 \delta_2 \delta_1 \|g_k\|. \quad (4.9)
 \end{aligned}$$

由 (4.9) 式, 假设 A6 和假设 A4 知

$$\frac{\|g_{k+1}\|}{\|s_k\|} \rightarrow 0, k \rightarrow \infty. \quad (4.10)$$

由 (4.7) 式知

$$\frac{\|g_{k+1}\|}{\|g_k\|} = \frac{\|g_{k+1}\|}{\|s_k\|} \cdot \frac{\|s_k\|}{\|g_k\|} \leq \frac{1}{w} \left(1 + \sqrt{\frac{W}{w}}\right) \cdot \frac{\|g_{k+1}\|}{\|s_k\|}. \quad (4.11)$$

由 (4.11) 和 (4.10) 式知

$$\lim_{k \rightarrow \infty} \frac{\|g_{k+1}\|}{\|g_k\|} = 0. \quad (4.12)$$

由 (4.2) 式和假设 A4 知

$$w \|x_{k+1} - x^*\| \leq \|g_{k+1}\|, \quad (4.13)$$

$$\|g_k\| \leq W \|x_k - x^*\|. \quad (4.14)$$

由 (4.13) 和 (4.14) 式知

$$0 \leq \frac{w \|x_{k+1} - x^*\|}{W \|x_k - x^*\|} \leq \frac{\|g_{k+1}\|}{\|g_k\|}. \quad (4.15)$$

由 (4.15) 和 (4.12) 式知

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.$$

故 $\{x_k\}$ 超线性收敛于 x^* .

5 数值试验

这一部分对算法 (CSATR) 进行了数值试验. 求解信赖域子问题 (1.6) 采用文 [12] 中的算法 3.1. B_k 和 b_k 的修正用文 [12] 公式进行修正. R -函数取为

$$R_\mu(r_k) = \begin{cases} \frac{2}{\pi} (M - 1 - \gamma_2) \arctan(r_k - \mu) + (1 + \gamma_2), & r_k \geq \mu; \\ (1 - \gamma_1 - \beta) \left(e^{r_k - \mu} + \frac{\beta}{1 - \gamma_1 - \beta} \right), & r_k < \mu. \end{cases}$$

算法 (CSATR) 中的参数选取如下: $\mu = 0.25, \gamma_1 = \gamma_2 = 0.15, \beta = 0.1, M = 5, \alpha = 0.5, \Delta_0 = 1, b_0 = 0, B_0 = I$ (单位矩阵). 测试问题取自文献 [14] 和文献 [15], 选取的初始点 x_0 均与原文献相同. 算法在 Matlab7.0 环境下编制程序运行. 实验结果如表 1. 在表 1 中, k, k_f, k_g 分别表示迭代次数, 函数值计算次数和梯度值计算次数. 终止准则为 $\varepsilon = 10^{-4}$.

表 1: 算法 (CSATR)

问题	维数	k	k_f	k_g
Pealty 1	2	23	47	24
	3	38	77	39
Extended Pealty	5	41	83	42
Extended PSC1	2	39	79	40
	4	41	83	42
Diagonal 1	10	67	135	68
	15	66	133	67
Diagonal 2	2	159	319	160
Diagonal 3	2	80	161	81
Extended Tridiagonal 2	2	21	43	22
Hager	2	83	167	84
Perturbed Quadratic	4	41	83	42
	10	41	83	42
Raydan 2	2	86	173	87
	5	91	183	92
Arwhead	2	21	43	22
Quadratic QF1	10	88	177	89
	15	88	177	89
Full Hessian FH2	2	136	273	137
	4	170	341	171

从由表 1 可以看出: 对一些函数, 迭代次数, 函数值计算次数及梯度值计算次数都是比较少的, 因此算法 (CSATR) 是一种有效的算法.

参 考 文 献

- [1] Schnabel R B, Eskow E. A new modified Cholesky factorization[J]. SIAM J. Sci. Stat. Comput., 1990, 11: 1136–1158.
- [2] Powell M J D. A new algorithm for unconstrained optimization[A]. Rosen J B, Mangasarian O L, Ritter K. Nonlinear programming[C]. New York: Academic Press, 1970: 31–66.
- [3] Zhang X S, Zhang J L, Liao L Z. An adaptive trust region method and its convergence[J]. Science in China (Series A), 2002, 45: 620–631.
- [4] 姚升保, 施保昌, 彭叶辉. 一类带线搜索的非单调信赖域算法 [J]. 数学杂志, 2003, 23(2): 290–294.
- [5] Hei L. A self-adaptive trust region algorithm[J]. J. Comput. Math., 2003, 21: 229–236.

- [6] Davidon D C. Conic approximations and collinear scaling for optimizers[J]. *SIAM J. Numer. Anal.*, 1980, 17(2): 268–281.
- [7] Sorenson D C. The Q -superlinear convergence of a collinear scaling algorithm for unconstrained optimization[J]. *SIAM J. Numer. Anal.*, 1980, 17: 84–114.
- [8] Ariyawansa K A. Deriving collinear scaling algorithms as extension of quasi-Newton methods and local convergence of DFP and BFGS-related collinear scaling algorithms[J]. *Math., Prog.*, 1990, 49: 23–48.
- [9] Sheng S. Interpolation by conic model for unconstrained optimization[J]. *Comput.*, 1995, 54: 83–98.
- [10] Zhu M F, Xue Y, Zhang F S. A quasi-Newton type trust region method based on the conic model[J]. *J. Chinese Uni.*, 1995, 17(1): 36–47.
- [11] Ni Q. Optimality condition for trust-region subproblems involving a conic model[J]. *SIAM J. Optim.*, 2005, 15: 826–837.
- [12] Fu J H, Sun W Y, Taimundo J B De Sampaio. An adaptive approach of conic trust region method for unconstrained optimization[J]. *J. Appl. Math. Comput.*, 2005, 19(1-2): 165–177.
- [13] 王希云, 王庆. 一种基于新锥模型的自适应信赖域算法 [J]. *应用数学*, 2010, 23(2): 307–312.
- [14] More J J, Burton S, Garbow, Hillstom K E. Testing unconstrained optimization software[J]. *ACM Trans. Math. Soft.*, 1981, 7(1): 17–41.
- [15] Neculai Andrei. An unconstrained optimization test functions collection[J]. *Advanced Model. Optim.*, 2008, 10(1): 147–161.

A SELF-ADAPIVE TRUST REGION METHOD AND ITS CONVERGENCE WITH THE CONIC MODEL

FENG Lin¹, DUAN Fu-jian²

(1.School of Math. and Finance, Chongqing University of Arts and Sciences, Chongqing 402160, China)

(2.School of Math. and Comput. Sci., Guilin University of Electronic Technology,
Guilin 541004, China)

Abstract: In this paper, we study a self-adaptive trust region algorithm based on the conic model for unconstrained optimization problems. A new self-adaptive trust region radius is produced under theoretical analysis. At each iterative, the trust region radius is updated at a variable rate, the information at the current point and the level vector information. We analyze the global convergence and Q -quadratic convergence of the new method. Numerical results are also presented to test the efficiency of the new method which extend the application and efficiency of self-adaptive trust region algorithms.

Keywords: unconstrained optimization; trust region method; conic model; self-adaptive; convergence

2010 MR Subject Classification: 90C30; 65K10