系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
29(4) (2009, 4), 470–483

基于简单二次函数模型的非单调信赖域算法*

孙清滢 段立宁 崔彬
(中国石油大学数学与计算科学学院，东营 257061)

王长钰
(曲阜师范大学 (日照校区) 运筹与管理学院，日照 276826)

摘要 基于简单二次函数模型，结合非单调技术，建立了一个新的求解无约束最优化问题的非单调信赖域算法，并证明了算法的全局收敛性及超线性收敛性. 数值例子表明算法是有效性的，适合求解大规模问题.

关键词 无约束最优化，非单调信赖域算法，超线性收敛，数值实验.

MR(2000) 主题分类号 90C30

1 引言

考虑无约束优化问题

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

其中 $f(x): R^n \to R^1$ 是一阶连续可微函数.

众所周知，信赖域算法 [1−9] 是求解问题 (1) 的重要算法，具有较强的全局收敛性质和较快的局部收敛速度. 信赖域子问题的求解是信赖域算法的一个关键问题，算法的工作量大部分集中在子问题的求解上. 通常的信赖域子问题基于如下二次模型

$$\begin{aligned} \min\quad & q_k(s) = f(x_k) + g_k^{\mathrm{T}} s + \frac{1}{2} s^{\mathrm{T}} B_k s, \\ \text{s.t.}\quad & \|s\| \le \triangle_k. \end{aligned} \tag{2}$$

其中 $s = x - x_k$, $g_k = \nabla f(x_k)$ 是目标函数 $f(x)$ 在当前点 x_k 的梯度. B_k 是目标函数 $f(x)$ 的 Hesse 矩阵或其近似，$\triangle_k \ge 0$ 是信赖域半径，$\|\cdot\|$ 为某一范数，通常采用 2- 范数.

* 国家自然科学基金 (10571106) 和中国石油大学博士基金 (Y040804) 资助项目.
收稿日期：2007-09-14, 收到修改稿日期：2008-07-08.

易见, 求解问题 (2) 之困难在于 B_k 是一个一般的实对称矩阵, 如果 B_k 是一个实对称正定对角矩阵, 则可以较方便地求出子问题 (2) 的最优解 s^*, 即若 $\|B_k^{-1}g_k\| \le \triangle_k$, 则 $s^* = B_k^{-1}g_k$; 若 $\|B_k^{-1}g_k\| > \triangle_k$, 求解 $\|\alpha B_k^{-1}g_k\| = \triangle_k$, 得 $\alpha = \frac{\triangle_k}{\|B_k^{-1}g_k\|}$, 则

$$s^* = -\alpha B_k^{-1}g_k = -\frac{\triangle_k}{\|B_k^{-1}g_k\|}B_k^{-1}g_k.$$

能否找到一个实对称正定对角矩阵 B_k 是目标函数 $f(x)$ 的 Hesse 矩阵或其近似呢？事实上, 结合时贞军 [10] 的稀疏对角拟牛顿方法, 可取 $B_k = \mathrm{diag}(b_k^1, b_k^2, \cdots, b_k^n)$ 满足

$$\min \|y_{k-1} - B_k s_{k-1}\|^2, \tag{3}$$

其中 $s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = g_k - g_{k-1}$. 同时为保证 B_k 是正定矩阵, 限制 b_k^i 在一定范围内取值, 即要求 $0 < \underline{L} \le b_k^i \le \overline{L}$, $\forall i = 1, 2, \cdots, n$. 为此需要求解如下问题

$$\min_{\underline{L} \le b_k^i \le \overline{L}, i=1,2,\cdots,n} \|y_{k-1} - B_k s_{k-1}\|^2, \tag{4}$$

即

$$\min_{\underline{L} \le b_k^i \le \overline{L}, i=1,2,\cdots,n} \sum_{i=1}^{n} (y_{k-1}^i - b_k^i s_{k-1}^i)^2. \tag{4$'$}$$

解之得， $\forall i = 1, 2, \cdots, n$, 当 $s_{k-1}^i \ne 0$ 时，若 $\underline{L} \le \frac{y_{k-1}^i}{s_{k-1}^i} \le \overline{L}$, 则取 $b_k^i = \frac{y_{k-1}^i}{s_{k-1}^i}$; 若 $\frac{y_{k-1}^i}{s_{k-1}^i} < \underline{L}$, 则取 $b_k^i = \underline{L}$; 若 $\frac{y_{k-1}^i}{s_{k-1}^i} > \overline{L}$ ，则取 $b_k^i = \overline{L}$. 当 $s_{k-1}^i = 0$ 时，则取 $b_k^i = \frac{\underline{L}+\overline{L}}{2}$.

本文取信赖域子问题模型为

$$\begin{aligned} \min \quad & q_k(s) = f(x_k) + g_k^{\mathrm{T}} s + \frac{1}{2} s^{\mathrm{T}} B_k s, \\ \text{s.t.} \quad & \|s\| \le \triangle_k. \end{aligned} \tag{5}$$

其中 $g_k = \nabla f(x_k)$, B_k 是实对称正定对角矩阵，满足 (4).

非单调搜索技术由于其有利于求解全局最优解和算法的快速收敛而受到许多最优化爱好者的青睐．早期的研究是 1980 年 Grippo Lamparello 和 Lueidi[11] 对牛顿算法提出的非单调线搜索技术，即步长 α_k 满足

$$f(x_k + \alpha_k d_k) \le \max_{0 \le j \le m_k} f(x_{k-j}) + \beta \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k, \tag{6}$$

其中 $\beta \in (0, 1)$, $0 \le m_k \le \min\{m_{k-1} + 1, M\}$, M 是给定的非负整数．易见，当 $m_k = 0$ 时，(6) 即为 Armijio 单调搜索．最近， Zhang H C, Willian W Hager[12] 改进了传统非单调线搜索技术中的参考值 $\max_{0 \le j \le m_k} f(x_{k-j})$, 提出了新的非单调线搜索技术

$$f(x_k + \alpha d_k) \le C_k + \beta \alpha \nabla f(x_k)^{\mathrm{T}} d_k, \tag{7}$$

其中

$$C_k = \begin{cases} f(x_k), & k = 0, \\ \dfrac{\eta_{k-1} Q_{k-1} C_{k-1} + f(x_k)}{Q_k}, & k \ge 1. \end{cases} \tag{8}$$

$$Q_k=\begin{cases}1, & k=0,\\ \eta_{k-1}Q_{k-1}+1, & k\ge 1.\end{cases} \tag{9}$$

这里 $\eta_{k-1}\in[\eta_{\min},\eta_{\max}]$, $\eta_{\min}\in[0,1)$ 和 $\eta_{\max}\in[\eta_{\min},1]$ 是两个参数. 结合 Zhang H C 等提出的非单调技术， Mo Jiangtao, Liu Chunyan, Yan Shicui[13] 建立求解无约束最优化问题 (1) 的信赖域算法，但算法仍然需要存储和计算一个 n 阶矩阵，无法求解大规模问题.

本文将基于简单二次模型 (5), 结合 Zhang H C 等提出的非单调技术，建立求解无约束最优化问题 (1) 的信赖域算法，并证明算法的全局收敛性和算法的超线性收敛速度. 用数值例子验证算法的有效性及适合大规模问题求解.

2 算 法

设子问题 (5) 的解为 s_k, 则目标函数在第 k 步的实际下降量为

$$\mathrm{Are}(s_k)=C_k-f(x_k+s_k), \tag{10}$$

模型 (5) 的预估下降量为

$$\mathrm{Pre}(s_k)=q_k(0)-q_k(s_k), \tag{11}$$

定义比值

$$\rho_k=\frac{\mathrm{Are}(s_k)}{\mathrm{Pre}(s_k)}=\frac{C_k-f(x_k+s_k)}{q_k(0)-q_k(s_k)}, \tag{12}$$

其中 C_k 由 (8), (9) 式确定. 下面给出非单调信赖域算法 (NTR).

给定初始点 $x_0\in R^n$, $\triangle_{\max}>0$, $\triangle_0>0$, $\mu\in(0,1)$, $0<c_1<c_2<1$, $c_3>1$, $B_0=I$, $\eta_{\min}\in[0,1)$, $\eta_{\max}\in[\eta_{\min},1]$, 令 $k:=0$, 转步骤 1.

步骤 1 计算 g_k, 如果 $\|g_k\|\le\varepsilon$, 则停， $x^*=x_k$, 否则，转步骤 2.

步骤 2 求解信赖域子问题 (5) 得解 s_k.

步骤 3 利用 (8), (9) 式计算 C_k, 利用 (10), (11), (12) 式计算 ρ_k.

步骤 4 令

$$x_{k+1}=\begin{cases}x_k+s_k, & \rho_k\ge\mu,\\ x_k, & \rho_k<\mu.\end{cases} \tag{13}$$

步骤 5 调整 $\triangle_{k+1}$ 如下: 如果 $\rho_k<\mu$, 取 $\triangle_{k+1}\in[c_1\|s_k\|,c_2\triangle_k]$; 如果 $\rho_k\ge\mu$, $\|s_k\|<\triangle_k$, 取 $\triangle_{k+1}=\triangle_k$; 如果 $\rho_k\ge\mu$, $\|s_k\|=\triangle_k$, 取 $\triangle_{k+1}\in[\triangle_k,c_3\triangle_k]$.

步骤 6 求解问题 (4) 得 B_{k+1}, 选取 $\eta_k\in[\eta_{\min},\eta_{\max}]$, 令 $k:=k+1$, 转步骤 1.

对算法 (NTR) 仿照文献 [2, 3] 的相关证明，可证明如下引理.

引理 1 设 s_k 是问题 (5) 的解，若 $g_k\ne 0$, 则

$$\triangle q_k(s_k)=q_k(0)-q_k(s_k)>0.$$

为得到算法的下降性条件，引入 Cauchy 点

$$s_k^c=-\tau_k\frac{\triangle_k}{\|g_k\|}g_k, \tag{14}$$

其中

$$\tau_k = \min\Big\{\frac{\|g_k\|^3}{\triangle_k g_k^{\mathrm{T}} B_k g_k}, 1\Big\}. \tag{15}$$

引理 2　对 Cauchy 点 s_k^c 满足

$$q_k(0) - q_k(s_k^c) \geq \frac{1}{2}\|g_k\| \min\Big\{\triangle_k, \frac{\|g_k\|}{\|B_k\|}\Big\}. \tag{16}$$

引理 3　设 s_k 是问题 (5) 的精确解，则

$$\mathrm{Pre}(s_k) \geq \frac{1}{2}\|g_k\| \min\Big\{\triangle_k, \frac{\|g_k\|}{\|B_k\|}\Big\}. \tag{17}$$

3　算法的全局收敛性

引理 4　设 $\{x_k\}$ 是由算法产生无穷迭代点列，则有

$$f_{k+1} \leq C_{k+1} \leq C_k, \quad \forall k = 1, 2, \cdots. \tag{18}$$

证　i) 对 $k \in I = \{k : \rho_k \geq \mu\}$, 由 $\rho_k \geq \mu$, 式 (12) 和引理 3 知

$$f_{k+1} \leq C_k - \frac{1}{2}\mu\|g_k\| \min\Big\{\triangle_k, \frac{\|g_k\|}{\|B_k\|}\Big\}. \tag{19}$$

再由 (8), (9) 式和 (19) 式知

$$\begin{aligned} C_{k+1} &= \frac{\eta_k Q_k C_k + f(x_{k+1})}{Q_{k+1}} \\ &\leq \frac{\eta_k Q_k C_k + C_k - \frac{1}{2}\mu\|g_k\| \min\{\triangle_k, \frac{\|g_k\|}{\|B_k\|}\}}{Q_{k+1}} \\ &= C_k - \frac{\frac{1}{2}\mu\|g_k\| \min\{\triangle_k, \frac{\|g_k\|}{\|B_k\|}\}}{Q_{k+1}}. \end{aligned} \tag{20}$$

再由 (8) 式知

$$C_{k+1} - C_k = \frac{f_{k+1} - C_{k+1}}{\eta_k Q_k}.$$

再由 (20) 式知

$$f_{k+1} \leq C_{k+1} \leq C_k, \quad \forall k \in I. \tag{21}$$

ii) 对 $k \in J = \{k : \rho_k < \mu\}$, 由算法步骤 4 知 $x_{k+1} = x_k$, 且 $f_{k+1} = f_k$. 首先证明 $f_{k+1} \leq C_{k+1}, \forall k \in J$. 分两种情况：

情况 1　$k-1 \in I$. 此时由 (21) 知 $f_k \leq C_k$. 再由 (8) 式，(9) 式和 $f_{k+1} = f_k$ 知

$$C_{k+1} = \frac{\eta_k Q_k f_k + f(x_{k+1})}{Q_{k+1}} = \frac{\eta_k Q_k f_{k+1} + f(x_{k+1})}{Q_{k+1}} = f(x_{k+1}). \tag{22}$$

情况 2　$k-1 \in J$, 此时定义

$$K=\{i: 1<i<k, k-i \in I\}.$$

如果 $K=\Phi$, 由算法步骤 4 知

$$f_0=f_{k-j}=f_{k+1}, \quad j=0,1,2,\cdots,k-1.$$

再由 (8) 式， (9) 式知

$$C_{k+1}=C_k=f_{k+1}. \tag{23}$$

如果 $K \neq \Phi$, 令 $m=\min\{i: i \in K\}$, 则有

$$f_{k-j}=f_k=f_{k+1}, \quad j=1,2,\cdots,m-1. \tag{24}$$

再由 (8) 式知

$$Q_k C_k=\eta_{k-1} Q_{k-1} C_{k-1}+f_k, \quad k \geq 1. \tag{25}$$

反复应用 (25) 式，得

$$\eta_k Q_k C_k+f_{k+1}=\prod_{i=0}^{m-1} \eta_{k-i} Q_{k-m+1} C_{k-m+1}+\sum_{j=0}^{m-2} \prod_{i=0}^{j} \eta_{k-i} f_{k-j}+f_{k+1}. \tag{26}$$

再由 K, m 的定义和式 (21), 有 $k-m \in I$, $C_{k-m+1} \geq f_{k-m+1}$. 再由 (24), (26) 知

$$\begin{aligned}
\eta_k Q_k C_k+f_{k+1} & \geq \prod_{i=0}^{m-1} \eta_{k-i} Q_{k-m+1} f_{k-m+1}+\sum_{j=0}^{m-2} \prod_{i=0}^{j} \eta_{k-i} f_{k-j}+f_{k+1} \\
& =\Big(\prod_{i=0}^{m-1} \eta_{k-i} Q_{k-m+1}+\sum_{j=0}^{m-2} \prod_{i=0}^{j} \eta_{k-i}+1\Big) f_{k+1} \\
& =Q_{k+1} f_{k+1}.
\end{aligned} \tag{27}$$

再由 (8) 式， (27) 式得

$$C_{k+1}=\frac{\eta_k Q_k f_k+f(x_{k+1})}{Q_{k+1}} \geq \frac{Q_{k+1} f_{k+1}}{Q_{k+1}}=f_{k+1}. \tag{28}$$

综合 (22), (23) 和 (28) 式知 $f_{k+1} \leq C_{k+1}, \forall k \in J$. 由于 C_{k+1} 是 f_{k+1} 和 C_k 的凸结合，因此 $C_{k+1} \leq C_k$. 故有 $f_{k+1} \leq C_{k+1} \leq C_k, \forall k \in J$. 因此 $\forall k$ 有 $f_{k+1} \leq C_{k+1} \leq C_k$.

引理 5　假设水平集 $\Omega_0=\{x | f(x) \leq f(x_0)\}$ 有界，则算法产生的点列 $\{x_k\} \subset \Omega_0$.

证　由引理 4 及 $C_0=f_0$ 易知结论成立.

算法中，如果 $x_{k+1}=x_k+d_k$, 则称 x_{k+1} 为一个成功的迭代点；如果 $x_{k+1}=x_k$, 则称 x_{k+1} 为一个非成功的迭代点.

引理 6　假设水平集 $\Omega_0=\{x | f(x) \leq f(x_0)\}$ 有界， $\{x_k\}$ 是由算法产生的无穷点列. 如果 $\forall k$ 有 $\|g_k\| \geq \delta$, $\delta \in(0,1)$ 为常数，则 $\forall k$, 存在非负整数 m 使得 x_{k+m+1} 是一个成功的迭代点.

证　(反证法) 假设存在 k_0 使得 $\forall m$ 都有 x_{k+m+1} 是非成功的迭代点，即

$$\rho_{k_0+m}<\mu, \quad m=0,1,2,\cdots. \tag{29}$$

由算法步骤 4 和步骤 5 知

$$x_{k_0+m+1}=x_k,\quad m=0,1,2,\cdots,\tag{30}$$

且

$$\triangle_{k_0+m+1}\to 0,\quad m\to\infty.\tag{31}$$

故对充分大的 m, 由 $f(x)$ 是连续可微函数，式 (30) 和 $\|s_k\|\le\triangle_k$ 及 Taylor 展开式知

$$\begin{aligned}&|f_k-f(x_k+s_{k+m})-(q_{k+m}(0)-q_{k+m}(s_{k+m}))|\\ =&\Big|\frac{1}{2}s_{k+m}^{\mathrm{T}}B_{k+m}s_{k+m}-\int_0^1[g(x_k+ts_{k+m})-g(x_k)]^{\mathrm{T}}s_{k+m}\mathrm{d}t\Big|\\ \le& O(\|B_{k+m}\|\triangle_{k+m}^2)+o(\triangle_{k+m}).\end{aligned}\tag{32}$$

因而对充分大的 m, 由引理 3, $\|g_k\|\ge\delta$, (30), (31), (32) 知

$$\Big|\frac{f_k-f(x_k+s_{k+m})}{q_{k+m}(0)-q_{k+m}(s_{k+m})}-1\Big|\le\frac{O(\|B_{k+m}\|\triangle_{k+m}^2)+o(\triangle_{k+m})}{\frac{1}{2}\delta\min\{\triangle_{k+m},\frac{\delta}{\|B_{k+m}\|}\}}.$$

再由 (31) 及 $\{B_k\}$ 的一致正定性知

$$\lim_{m\to\infty}\frac{f_k-f(x_k+s_{k+m})}{q_{k+m}(0)-q_{k+m}(s_{k+m})}=1.\tag{33}$$

另外，由 (12), 引理 4 和 (29) 知

$$\rho_{k+m}=\frac{C_{k+m}-f(x_{k+m}+s_{k+m})}{q_{k+m}(0)-q_{k+m}(s_{k+m})}\ge\frac{f_k-f(x_k+s_{k+m})}{q_{k+m}(0)-q_{k+m}(s_{k+m})}.\tag{34}$$

因此当 m 充分大时，由 (33) 和 (34) 及 $\mu\in(0,1)$ 知 $\rho_{k+m}\ge\mu$, 此与 (29) 式矛盾.

定理 1 假设水平集 $\Omega_0=\{x|f(x)\le f(x_0)\}$ 有界，$\{x_k\}$ 是由算法产生的无穷点列，则 $\liminf\limits_{k\to\infty}\|g_k\|=0$.

证 (反证法) 假设结论不成立，则 $\exists$ 常数 $\delta>0$, $\delta\in(0,1)$, 使得

$$\|g_k\|\ge\delta,\quad\forall k.\tag{35}$$

下证

$$\lim_{k\to\infty}\triangle_k=0.\tag{36}$$

首先证明

$$\lim_{k\to\infty,k\in I}\triangle_k=0.\tag{37}$$

由引理 6 知，I 是一个无穷子集，对 $k\in I$, 由 (20) 和 (35) 知

$$C_{k+1}\le C_k-\frac{\frac{1}{2}\mu\delta\min\{\triangle_k,\frac{\delta}{\|B_k\|}\}}{Q_{k+1}}.\tag{38}$$

由引理 4 知 $f_{k+1}\le C_{k+1}$, $\forall k$, 且 $\{C_k\}$ 是一个单调不增序列. 由假设水平集 $\Omega_0=\{x|f(x)\le f(x_0)\}$ 有界，引理 5 和 f 的连续性知，$\{f\}$ 是有下界的，因而 $\{C_k|k\in I\}$ 是收敛的，因此

$$0\le\lim_{k\to\infty,k\in I}\frac{\frac{1}{2}\mu\delta\min\{\triangle_k,\frac{\delta}{\|B_k\|}\}}{Q_{k+1}}\le\lim_{k\to\infty,k\in I}(C_k-C_{k+1})=0.$$

故

$$\lim_{k\to\infty,k\in I}\frac{\min\{\triangle_k,\frac{\delta}{\|B_k\|}\}}{Q_{k+1}}=0. \tag{39}$$

由 $\eta_{\max}\in[0,1)$ 和 $\eta_k\in[\eta_{\min},\eta_{\max}]$, 反复应用 (9) 式知

$$\begin{aligned}Q_{k+1}&=1+\sum_{i=0}^{k}\prod_{m=0}^{i}\eta_{k-m}\le 1+\sum_{i=0}^{k}\eta_{\max}^{i+1}\le\sum_{i=0}^{\infty}\eta_{\max}^{i}\\&=\frac{1}{1-\eta_{\max}}.\end{aligned} \tag{40}$$

由 (39) 和 (40) 及 $\{B_k\}$ 的一致正定性知 (37) 成立.

进一步证明 $\lim\limits_{k\to\infty,k\in J}\triangle_k=0$. 如果 J 是有限集，则由 (37) 可推知 (36) 成立. 如果 J 是无限集，且 $k_1=\{i_k|k=1,2,\cdots\}$ 是 J 中的无限子集满足

$$i_k=\min\{j|j\in J\},\quad i_{k+1}=\min\{j\in J|j-1\in I,j>i_k\},\quad\forall k\ge 1.$$

对 $k\ge 1$, 由 i_k 的定义知 $i_k-1\in I$, 由算法步骤 5 知

$$\triangle_{i_k}\le c_3\triangle_{i_k-1}. \tag{41}$$

再由 k_1 和 i_k 的定义知，$\exists$ 非负整数 l_k 使得

$$i_k+l_k<i_{k+1}-1,\quad i_k+l\in J,\quad\forall\, l=0,1,2,\cdots,l_k.$$

令 $l_k^{'}=\max\{l_k\}$, 再由算法步骤 5 知

$$\triangle_{i_k+l}\ge\triangle_{i_k+l+1},\quad\forall l=0,1,2,\cdots,l_k^{'}. \tag{42}$$

由 (37) 知，$\triangle_{i_k-1}\to 0(k\to\infty)$. 再由 (41) 和 (42) 知

$$\lim_{k\to\infty,k\in J}\triangle_k=0. \tag{43}$$

因此由 (37) 和 (43) 知 (36) 成立. 下面证明定理.

由 f 是连续可微函数，$\|s_k\|\le\triangle_k$, $\forall k$, 式 (36) 及 Taylor 展开式知

$$\begin{aligned}&|f_k-f(x_k+s_k)-(q_k(0)-q_k(s_k))|\\&=\Big|\frac{1}{2}s_k^{\mathrm T}B_ks_k-\int_0^1[g(x_k+ts_k)-g(x_k)]^{\mathrm T}s_k\mathrm{d}t\Big|\\&\le O(\|B_k\|\triangle_k^2)+o(\triangle_k).\end{aligned} \tag{44}$$

由引理 3, (35), (44) 知

$$\Big|\frac{f_k-f(x_k+s_k)}{q_k(0)-q_k(s_k)}-1\Big|\le\frac{O(\|B_k\|\triangle_k^2)+o(\triangle_k)}{\frac{1}{2}\delta\min\{\triangle_k,\frac{\delta}{\|B_k\|}\}}.$$

令 $k\to\infty$, 再由 (36) 式知

$$\lim_{k\to\infty}\frac{f_k-f(x_k+s_k)}{q_k(0)-q_k(s_k)}=1. \tag{45}$$

由 (12) 及引理 4 知

$$\rho_k = \frac{C_k - f(x_k + s_k)}{q_k(0) - q_k(s_k)} \ge \frac{f_k - f(x_k + s_k)}{q_k(0) - q_k(s_k)}. \tag{46}$$

这样，当 k 充分大时，由 (45), (46) 及 $\mu \in (0,1)$ 知有 $\rho_k \ge \mu$. 再由算法步骤 5 知，对任意充分大的 k 有 $\triangle_{k+1} \ge \triangle_k$, 此与 (36) 矛盾．证毕．

4 超线性收敛

假设 1 $f(x)$ 是二阶连续可微函数，序列 $\{x_k\}$ 收敛于 x^*, 且 $\nabla^2 f(x^*)$ 是正定矩阵．

假设 2 算法产生的序列 $\{x_k\}$ 及 $\{B_k\}$ 满足

$$\lim_{k\to\infty} \frac{\|(B_k - \nabla^2 f(x_k))s_k\|}{\|s_k\|} = 0. \tag{47}$$

定理 2 设 $\{x_k\}$ 是由算法产生的序列，算法中满足假设 1, 2, 则序列 $\{x_k\}$ 超线性收敛于 x^*.

证 由假设条件 1 知， $\exists M > m > 0$ 使得 $\forall x \in \Omega_1 = \{x | \|x - x^*\| \le \triangle\}$, $\triangle$ 为一个很小的常数

$$m\|z\|^2 \le z^{\mathrm T} \nabla^2 f(x) z \le M\|z\|^2, \quad \forall z \in R^n. \tag{48}$$

且存在充分大的正整数 k_0, 使得 $\forall k \ge k_0$ 有 $x_k \in \Omega_1$. 由中值定理知， $\forall k \ge k_0$ 有

$$\frac{1}{2} m\|x_k - x^*\|^2 \le f_k - f(x^*) \le \frac{1}{2} M\|x_k - x^*\|^2, \tag{49}$$

且

$$m\|x_k - x^*\|^2 \le \|g_k\| \le M\|x_k - x^*\|^2. \tag{50}$$

由 (47) 和 (48) 知，对充分大的 $k > k_0$, B_k 满足

$$\begin{gathered} \frac{1}{2} m\|s_k\|^2 \le s_k^{\mathrm T} B_k s_k \le 2M\|s_k\|^2, \\ \|B_k\| \le 2M. \end{gathered} \tag{51}$$

下证对充分大的 $k > k_0$, $\exists$ 常数 c_0 使得

$$q_k(0) - q_k(s_k) \ge c_0 \|s_k\|^2. \tag{52}$$

事实上，由 (47) 知，对充分大的 $k > k_0$ 有

$$\|(B_k - \nabla^2 f(x_k))s_k\| \le \frac{1}{2} m\|s_k\|. \tag{53}$$

由 (5), (53) 知

$$\begin{aligned} 0 \le q_k(0) - q_k(s_k) &= -g_k^{\mathrm T} s_k - \frac{1}{2} s_k^{\mathrm T} B_k s_k \\ &\le -g_k^{\mathrm T} s_k - \frac{1}{2} s_k^{\mathrm T} \nabla^2 f(x) s_k + \frac{1}{4} m\|s_k\|^2 \\ &\le \|g_k\| \cdot \|s_k\| - \frac{1}{4} m\|s_k\|^2. \end{aligned}$$

由此可得

$$\|g_k\| > \frac{1}{4}m\|s_k\|. \tag{54}$$

由 (51), (54) 和引理 3 及 $m < M$ 知

$$\begin{aligned}
&q_k(0) - q_k(s_k)\\
\geq\ &\frac{1}{2}\|g_k\| \min\left\{\triangle_k, \frac{\|g_k\|}{\|B_k\|}\right\}\\
\geq\ &\frac{1}{2}\cdot\frac{1}{4}\cdot m\|s_k\| \min\left\{\|s_k\|, \frac{m\|s_k\|}{4\|B_k\|}\right\}\\
\geq\ &\frac{1}{8}m\|s_k\|^2 \min\left\{1, \frac{m}{8M}\right\} \geq \frac{1}{64M}\cdot m^2\|s_k\|^2.
\end{aligned} \tag{55}$$

取 $c_0 = \frac{1}{64M}m^2 > 0$, 故 (52) 成立．下证，当 k 充分大时 $\rho_k \geq \mu$.

由中值定理知

$$\begin{aligned}
&f_k - f(x_k + s_k) - (q_k(0) - q_k(s_k))\\
=\ &\frac{1}{2}s_k^{\mathrm{T}}(B_k - \nabla^2 f(x_k))s_k + \frac{1}{2}s_k^{\mathrm{T}}(\nabla^2 f(x_k) - \nabla^2 f(x_k + \zeta_k s_k))s_k,
\end{aligned}$$

其中 $\zeta_k \in (0,1)$. 由 $\{x_k\}$ 是收敛的知 $s_k \to 0(k\to\infty)$, 再由 $\nabla^2 f(x)$ 的连续性及 (47) 式知

$$f_k - f(x_k + s_k) - (q_k(0) - q_k(s_k)) = o(\|s_k\|^2). \tag{56}$$

再由 (52), (56) 知

$$\begin{aligned}
&\left|\frac{f_k - f(x_k + s_k)}{q_k(0) - q_k(s_k)} - 1\right|\\
=\ &\left|\frac{f_k - f(x_k + s_k) - q_k(0) + q_k(s_k)}{q_k(0) - q_k(s_k)}\right| \leq \frac{o(\|s_k\|^2)}{c_0\|s_k\|^2}.
\end{aligned} \tag{57}$$

因此

$$\lim_{k\to\infty} \frac{f_k - f(x_k + s_k)}{q_k(0) - q_k(s_k)} = 1. \tag{58}$$

再由 (12), 引理 4 知

$$\rho_k = \frac{C_k - f(x_k + s_k)}{q_k(0) - q_k(s_k)} \geq \frac{f_k - f(x_k + s_k)}{q_k(0) - q_k(s_k)}. \tag{59}$$

由 (58), (59) 知，当 k 充分大时有 $\rho_k \geq \mu$. 因此由算法知，对充分大的 k 有信赖域半径 $\triangle_k$ 是增大的，即对充分大的 k,

$$\triangle_k \geq \alpha. \tag{60}$$

但由假设 1 和定理 1 知 $\lim\limits_{k\to\infty}\|g_k\| = 0$, 因此再由算法的构造知 $\{B_k\}$ 一致正定，因此有 $\lim\limits_{k\to\infty}\|B_k^{-1}g_k\| = 0$, 即

$$\lim_{k\to\infty}\|s_k\| = 0.$$

再由中值定理知

$$g_{k+1} = g_k + B_k s_k + (\nabla^2 f(x_k) - B_k)s_k + \int_0^1 [\nabla^2 f(x_k + ts_k) - \nabla^2 f(x_k)]s_k \mathrm{d}t$$
$$= (\nabla^2 f(x_k) - B_k)s_k + \int_0^1 [\nabla^2 f(x_k + ts_k) - \nabla^2 f(x_k)]s_k \mathrm{d}t.$$

再由 $\nabla^2 f(x)$ 的连续性知 $\|g_{k+1}\| \le \|(\nabla^2 f(x_k) - B_k)s_k\| + o(\|s_k\|)$. 即

$$\frac{\|g_{k+1}\|}{\|s_k\|} \le \frac{\|(\nabla^2 f(x_k) - B_k)s_k\|}{\|s_k\|} + \frac{o(\|s_k\|)}{\|s_k\|}. \tag{61}$$

再由 (47), (61) 知

$$\lim_{k\to\infty} \frac{\|g_{k+1}\|}{\|s_k\|} = 0. \tag{62}$$

再由 (54) 知

$$\lim_{k\to\infty} \frac{\|g_{k+1}\|}{\|g_k\|} = 0.$$

由 (50) 知

$$0 \le \frac{m\|x_{k+1} - x^*\|}{M\|x_k - x^*\|} \le \frac{\|g_{k+1}\|}{\|g_k\|}. \tag{63}$$

令 $k \to \infty$ 得

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

故 $\{x_k\}$ 超线性收敛于 x^*.

注 1　由定理 2 可知，新算法在一定条件下具有超线性收敛特征，特别是对变量可分离的目标函数和 Hesse 矩阵稀疏的目标函数具有较快的收敛速度.

5　数值试验

本节选择了文献 [14, 15] 的几个算例，利用 matlab 编制程序在 PIII.933 机器上对本文算法进行数值试验. 如果算法中 B_k 用 DFP 公式或 BFGS 公式校正，分别记为 DFPTR 和 BFGSTR. 算法中取 $\triangle_0 = 0.1$, $\triangle_{\max} = 2.8$, $\mu = 0.1$, $c_1 = 0.26$, $c_2 = 0.63$, $c_3 = 1.91$, $\eta_{\min} = 0.19$, $\eta_{\max} = 0.89$. 以下在精度要求 $\|\nabla f(x_k)\| \le 10^{-3}$ 下分别给出计算结果. 如果时间大于 600 秒或者计算不出结果 (0 做除数), 则用 $*****$ 表示. 表 1 到表 5 的每个单元格中从上到下 5 个数值分别表示维数 n=100, 1000, 5000, 10000, 20000 的相应结果.

例 1[15]

$$f(x) = \sum_{i=1}^n f_i(x)^2, \quad f_{2i-1}(x) = (x_{2i} - x_{i-1}^2), \quad f_{2i}(x) = 1 - x_{2i-1},$$

初始点 $x_0 = (-1.2, 1, -1.2, 1, \cdots, -1.2, 1)^{\mathrm{T}}$, 最优点 $x_{\mathrm{opt}} = (1, 1, \cdots, 1)^{\mathrm{T}}$, 最优值 $f(x_{\mathrm{opt}}) = 0$, 取 $\underline{L} = 0.598$, $\overline{L} = 112$, 数值结果见表 1.

表 1 例 1 的数值结果

算法	迭代次数	迭代时间	最优值
NTR	47	0.0780	$4.2086e-007$
	57	0.2350	$4.3170e-007$
	62	0.6880	$4.3280e-008$
	63	2.3130	$5.7154e-007$
	63	5.7500	$5.6412e-007$
BFGSTR	50	0.8280	$2.9983e-009$
	82	512.6870	$4.8686e-007$
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
DFPTR	70	1.9690	$2.4865e-007$
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****

例 2[15] (扩展 Powell 测试函数)

$$f(x)=\sum_{i=1}^{\frac{n}{4}}[(x_{4i-1}+10x_{4i-2})^2+5(x_{4i-1}-x_{4i})^2+(x_{4i-2}-2x_{4i-1})^2+10(x_{4i-3}-x_{4i})^4],$$

初始点 $x_0=(3,-1,0,3,\cdots,3,-1,0,3)^{\mathrm{T}}$, 最优点 $x_{\mathrm{opt}}=(0,0,\cdots,0)^{\mathrm{T}}$, 最优值 $f(x_{\mathrm{opt}})=0$, 取 $\underline{L}=0.396$, $\overline{L}=371.3$, 数值结果见表 2.

表 2 例 2 的数值结果

算法	迭代次数	迭代时间	最优值
NTR	84	0.1400	$1.7397e-008$
	222	1.2340	$2.6836e-005$
	106	2.5940	$1.2044e-006$
	357	18.5780	$6.8079e-005$
	110	11.5160	$9.3101e-007$
BFGSTR	179	2.2970	$8.4473e-006$
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
DFPTR	1076	19.7970	$-1.9609e-008$
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****

例 3[15] (扩展 Dixon 测试函数)

$$f(x)=\sum_{i=1}^{\frac{n}{10}}\left[(1-x_{10i-9})^2+(1-x_{10i})^2+\sum_{j=10i-9}^{10i-1}(x_j^2-x_{j+1})^2\right],$$

初始值 $x_0=(-2,-2,\cdots,-2)^{\mathrm{T}}$, 最优点 $x_{\mathrm{opt}}=(1,1,\cdots,1)^{\mathrm{T}}$, 最优值 $f(x_{\mathrm{opt}})=0$, 取 $\underline{L}=0.598$, $\overline{L}=381.5$, 数值结果见表 3.

表 3 例 3 的数值结果

算法	迭代次数	迭代时间	最优值
NTR	100	0.1400	$4.7852e-008$
	123	0.5150	$3.6226e-008$
	128	2.4680	$6.5261e-008$
	669	26.1250	$3.5243e-009$
	131	9.7190	$3.4885e-008$
BFGSTR	44	0.6090	$1.6720e-008$
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
DFPTR	52	1.4840	$-6.1871e-008$
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****

例 4[14] (三角函数)

$$f(x)=\sum_{i=1}^{n}f_i(x)^2,\quad f_i(x)=n-\sum_{j=1}^{n}\cos x_j+i(1-\cos x_i)-\sin x_i,\quad i=1,2,\cdots,n,$$

初始值 $x_0=(\frac{1}{n},\frac{1}{n},\cdots,\frac{1}{n})^{\mathrm{T}}$, 最优值 $f(x_{\mathrm{opt}})=0$, 取 $\underline{L}=0.598$, $\overline{L}=1000$, 数值结果见表 4.

表 4 例 4 的数值结果

算法	迭代次数	迭代时间	最优值
NTR	87	0.1400	$1.7526e-006$
	29	0.1710	$3.3194e-007$
	21	0.6720	$9.7793e-008$
	21	1.2650	$5.4618e-008$
	19	2.1100	$3.5654e-008$
BFGSTR	15	0.2180	$4.7707e-006$
	8	39.4530	$2.2658e-006$
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
DFPTR	29	0.6250	$-1.3396e-007$
	11	83.8440	$-5.6193e-007$
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****

例 5[14] (Broyden 三对角函数)

$$f(x)=\sum_{i=1}^{n} f_i(x)^2,\quad f_i(x)=(3-2x_i)x_i-x_{i-1}-2x_{i+1}+1,\quad i=1,2,\cdots,n,$$

初始值 $x_0=(-1,-1,\cdots,-1)^{\mathrm{T}}$, 最优值 $f(x_{\mathrm{opt}})=0$, 取 $\underline{L}=0.801$, $\overline{L}=0.8254$, 数值结果见表 5.

表 5 例 5 的数值结果

算法	迭代次数	迭代时间	最优值
NTR	68	0.2960	$2.3254e-009$
	65	0.5630	$1.4474e-005$
	58	2.8910	$1.2247e-004$
	86	17.9220	$7.9526e-010$
	107	52.8440	$1.6192e-009$
BFGSTR	100	1.7660	$5.4668e-009$
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
DFPTR	101	2.9690	$-5.6835e-009$
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****
	*****	*****	*****

计算结果表明新算法 (NTR) 不受问题规模的限制，均能求得问题的最优解．另外在最优值 f^* 的计算精度上，新算法精度高．在维数上，尽管维数成倍增长，但迭代次数和 CPU 时间没有成倍增长．因此算法 (NTR) 是有效的，适合求解大规模问题．

参 考 文 献

[1] Powell M J D. On the global convergence of trust region algorithms for unconstrained optimization. *Math. Prog.*, 1984, **29**: 297–303.

[2] Powell M J D. Convergence Properties of a Class of Minimization Algorithms. Nonlinear Programming, Academic Press: New York, 1975.

[3] Conn A R, Gould N I M and Toint Ph L. Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM J. Numer. Anal.*, 1988, **25**: 433–460.

[4] Steihaug T. The conjugate gradient method and trust region in large scale optimization. *SIAM Journal on Numerical Analysis*, 1983, **20**: 626–637.

[5] 袁亚湘．信赖域方法的收敛性．计算数学， 1994, **3**: 334–346.

[6] 柯小伍，韩继业．一类新的信赖域算法的全局收敛性．应用数学学报， 1995, **18**(4): 608–615.

[7] 李正锋，邓乃扬．一类新的非单调信赖域算法及其收敛性．应用数学学报， 1999, **22**(3): 457–465.

[8] Deng N Y, Xiao Y, Zhou F J. Nonmonotonic trust-region algorithms. *Journal of Optimization Theory and Applications*, 1993, **26**: 259–285.
[9] Zhang Xiangsun, Zhang Juliang, Liao Lizhi. An adaptive trust region method and its convergence. *Science in China* (*Series A*), 2002, **45**(5): 620–631.
[10] 时贞军，孙国．无约束优化问题的对角稀疏拟牛顿算法．系统科学与数学， 2006, **26**(1): 101–112.
[11] Grippo L, Lampariello F and Lucidi S. A nonmonotone line search technique for newton's method. *SIAM J. Numer. Anal.*, 1986, **23**(4): 707–716.
[12] Zhang H C, Hager William W. A nonmonotone line search technique and its application to unconstrained optimization. *SIAM. J. Optim.*, 2004, **14**(4): 1043–1056.
[13] Mo Jiangtao, Liu Chunyan, Yan Shicui. A nonmonotone trust region method based on nonincreasing technique of weighted average of the successive function values. *Journal of Computational and Applied Mathematics*, 2006, **209**(1): 97–108.
[14] Shi Z J, Shen J. New inexact line search method for unconstrained optimization. *Journal of Optimization Theory and Applications*, 2005, **127**(2): 425–445.
[15] Touati-Ahmed D, Storey C. Efficient hybrid conjugate gradient techniques. *Journal of Optimization Theory and Applications*, 1990, **64**(2): 379–397.

A NON-MONOTONE TRUST REGION ALGORITHM WITH SIMPLE QUADRATIC MODELS

SUN Qingying　DUAN Lining　CUI Bin

(*School of Mathematics and Computational Sciences, China University of Petroleum, Dongying* 257061)

WANG Changyu

(*College of Operations Research and Management, Qufu Normal University* (*Rizhao Campus*), *Rizhao* 276826)

Abstract A new non-monotone trust region algorithm with simple quadratic models is proposed. Under certain conditions, the global and super-linear convergence properties of this new method are proved. Numerical results show that the new algorithm is efficient, and attractive for large-scale optimization problems.

Key words Unconstrained optimization, non-monotone trust region method, super-linear convergence, numerical experiment.