第 33 卷　第 2 期　　西南师范大学学报(自然科学版)　　2008 年 4 月
Vol. 33　No. 2　　Journal of Southwest China Normal University (Natural Science Edition)　　Apr.　2008

文章编号：1000－5471(2008)02－0001－05

非单调无约束最优化信赖域方法①

杨 正 豪

南京师范大学 数学与计算机科学学院，江苏 南京 210097

摘要： 给出了解无约束最优化非单调信赖域方法，该方法允许目标函数值在某些步上升，而保持其全局收敛性. 数值试验表明，非单调信赖域方法优于通常的信赖域方法.

关　键　词： 无约束最优化；信赖域方法；非单调方法；收敛性

中图分类号： O221.2　　**文献标识码：** A

对于无约束最优化问题

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

其中 $f(x)$ 为连续可微函数. 为了保证算法的总体收敛性，一般采用保持目标函数值单调下降线性搜索与信赖域两种最基本的方法[1-4]. 但这两种方法往往会使算法收敛速度减慢，尤其在出现弯曲峡谷(curved narrow valleys)的地方，导致搜索步长非常小，甚至出现“锯齿”现象. 本文给出非单调信赖域法(简称 NTR 法)，将文献[5]中非单调步长接受准则应用到文献[6]中，允许目标函数值在某些步上升，而保持其全局收敛性. 近年来，许多作者将非单调技术与其它技术相结合，广泛用来解各种光滑与非光滑优化问题[7-10]. 关于最优化问题非单调线性搜索方法见文献[5,11,12]. 如无特别说明，本文的向量范数为任意范数，矩阵范数为与其对应向量范数相容的矩阵范数.

1　NTR 算法模型

为了便于对照，先给出通常的解无约束最优化 UTR 方法(即通常信赖域方法)的算法模型. 设在第 k 次迭代 UTR 法的模型问题是

$$\phi_k(s) = g_k^T s + \frac{1}{2} s^T B_k s \tag{1}$$

其中：$g_k = g(x_k) \in R^n$ 是 $f(x)$ 的梯度 $\nabla f(x_k)$ 近似，$B_k \in R^{n\times n}$ 是 $f(x)$ 的近似 Hessian 对称矩阵，$s = x - x_k \in R^n$. 为方便起见，记 $f_k = f(x_k)$. 下面是 UTR 法的算法具体模型.

算法 1　UTR 法

步骤 1　给出初始点 $x_0 \in R^n$，$0 < \bar{\Delta} \in R^1$，$0 < \eta < \mu < 1$，$0 < \gamma_1 < \gamma_2 < 1 < \gamma_3$，$0 < \Delta_0 \leqslant \bar{\Delta}$，$\epsilon > 0$.

步骤 2　令 $k = 0$；计算 $f_0 = f(x_0)$.

步骤 3　计算 g_k，如果 $\| g_k \| \leqslant \epsilon$，则停，令 $x^* = x_k$；否则计算 B_k.

步骤 4　决定下面子问题的一个近似解 s_k

$$\min\{\phi_k(s):\ \| s \| \leqslant \Delta_k\}$$

其中 $\phi_k(s)$ 由(1)定义

步骤 5　计算 $f_{k+1} = f(x_k + s_k)$，$\rho_k = (f_k - f_{k+1})/(-\phi_k(s_k))$.

步骤 6　如果 $\rho_k \geqslant \mu$，令 $x_{k+1} = x_k + s_k$，转步骤 7；如果 $\eta < \rho_k < \mu$，令 $\Delta_k = \gamma_2 \Delta_k$，转步骤 4；如果 ρ_k

① 收稿日期：2007－09－18
作者简介：杨正豪(1971－)，男，江苏盱眙人，博士研究生，主要从事最优化理论及应用的教学与研究.

$\leqslant \eta$，令 $\Delta_k = \gamma_1 \Delta_k$，转步骤 4.

步骤 7　选取 Δ_{k+1} 使其满足

$$\Delta_{k+1} = \min\{\gamma_3 \Delta_k, \overline{\Delta}\}$$

令 $k = k+1$，转步骤 3.

注　在算法 1 中，步长 s_k 被接受当且仅当 $\rho_k \geqslant \mu$，因此，算法在每次迭代都保证目标函数值单调下降. 松弛对步长 s_k 的接受条件，就可得到非单调信赖域法(NTR).

算法 2　NTR 法

步骤 1　给出初始点 $x_0 \in R^n$，$0 < \overline{\Delta} \in R^1$，$0 < \eta < \mu < 1$，$0 < \gamma_1 < \gamma_2 < 1 < \gamma_3$，$0 < \Delta_0 \leqslant \overline{\Delta}$，$\epsilon > 0$，一个非负整数 M.

步骤 2　令 $k = 0$，$m(k) = 0$，计算 $f_0 = f(x_0)$.

步骤 3　计算 g_k，如果 $\| g_k \| \leqslant \epsilon$，则停，令 $x^* = x_k$；否则计算 B_k 与

$$f_{l(k)} = f(x_{l(k)}) = \max_{0 \leqslant j \leqslant m(k)} \{f_{k-j}\} \tag{2}$$

步骤 4　决定下面子问题的一个近似解 s_k

$$\min \{\phi_k(s): \| s \| \leqslant \Delta_k\} \tag{3}$$

其中 $\phi_k(s)$ 由(1) 定义.

步骤 5　计算 $f_{k+1} = f(x_k + s_k)$ 和

$$\rho_k = (f_{l(k)} - f_{k+1})/(-\phi_k(s_k)) \tag{4}$$

步骤 6　如果 $\rho_k \geqslant \mu$，令 $x_{k+1} = x_k + s_k$，转步骤 7；如果 $\eta < \rho_k < \mu$，令 $\Delta_k = \gamma_2 \Delta_k$，转步骤 4；如果 $\rho_k \leqslant \eta$，令 $\Delta_k = \gamma_1 \Delta_k$，转步骤 4.

步骤 7　选取 Δ_{k+1} 使其满足

$$\Delta_{k+1} = \min\{\gamma_3 \Delta_k, \overline{\Delta}\}$$

令 $k = k+1$，$m(k) = \min\{m(k-1)+1, M\}$，转步骤 3.

注 2　算法 1 与算法 2 主要区别是对步长 s_k 的接受准则不同，算法 2 接受步长 s_k 是当 $f(x_k + s_k)$ 与 $f(x_{l(k)})$ 比较有一定下降. 因此，接受准则 $\rho_k \geqslant \mu$ 并不意味着 $f_{k+1} < f_k$，即$\{f_k\}$ 不是单调下降的.

2　收敛性

为了讨论算法 2 的收敛性，设在算法 2 中的参数 $\epsilon = 0$，并设水平集

$$\Omega_0 = \{x \in R^n \mid f(x) \leqslant f(x_0)\}$$

是紧的. 在这个假设之下，在 Ω_0 中，$f(x)$ 是有下界的，且梯度函数 $\nabla f(x)$ 是一致连续的. 另外设子问题(3) 的近似解 s_k 与迭代矩阵 B_k 满足下面条件：

(Ⅰ) 存在两个正常数 c_1 和 c 满足

$$-\phi_k(s_k) \geqslant c_1 \| g_k \| \min\{\min\{\Delta_k, c\| g_k \|\}, \| g_k \| / \| B_k \|\}, k = 1,2,\cdots$$

(Ⅱ) 存在一个正常数 $\bar{c}$ 满足

$$\| s_k \| \leqslant \bar{c} \| g_k \|, k = 1,2,\cdots$$

(Ⅲ) 存在一个正常数 σ_1 满足

$$\| B_k \| \leqslant \sigma_1, k = 1,2,\cdots$$

考虑$\{x_k\}$ 极限点的性质，主要是研究 $\| g_k \|$ 的极限性质. 为此，先给出以下引理：

引理 1　假设条件(Ⅰ)成立，则$\{x_k\} \subset \Omega_0$.

证　用数学归纳法证明. 当 $k = 0$ 时，结论显然成立；假设 $k \leqslant m$ 时结论成立，则 $k = m+1$ 时，由算法 2 有

$$f_{m+1} \leqslant f_{l(m)} + \mu\phi_m(s_m) < f_{l(m)}$$

由 $l(m) \leqslant m$ 知，$x_{m+1} \in \Omega_0$. 从而引理成立.

引理 2　设 $f(x)$ 是连续的，且条件(Ⅰ)成立，则$\{f_{l(k)}\}$ 是单调减少且收敛的，其中$\{f_{l(k)}\}$ 由(2) 决定.

证　由 $m(k+1) \leqslant m(k)+1$ 知

$$f_{l(k+1)} = \max_{0 \leqslant j \leqslant m(k+1)} \{f_{k+1-j}\} \leqslant \max_{0 \leqslant j \leqslant m(k)+1} \{f_{k+1-j}\}$$
$$= \max\{f_{l(k)}, f_{k+1}\}$$

$$\leqslant f_{l(k)}$$

即$\{f_{l(k)}\}$单调减少．再由Ω_0是紧的知$\{f_{l(k)}\}$是收敛的．

引理 3　设$f(x)$是连续可微的，并且条件(Ⅰ)，(Ⅱ)，(Ⅲ)成立．那么

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

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

证　证明类似于文献[6]中的引理 3.3 证明．

定理 1　设$f(x)$是连续可微的，$\nabla f(x)$一致连续，如果条件(Ⅰ)和(Ⅲ)成立，且当$\lim\limits_{k\to\infty}\|s_k\|=0$时，有$\lim\limits_{k\to\infty}\|\nabla f(x_k)-g_k\|=0$，则有

$$\underline{\lim_{k\to\infty}}\|g(x_k)\|=0$$

证　用反证法．假设$\underline{\lim\limits_{k\to\infty}}\|g(x_k)\|\neq 0$，即存在$\epsilon_0>0$和充分大的正整数$K(>M)$，对任意$k>K$，有

$$\|g_k\|\geqslant\epsilon_0$$

于是由类似于文献[6]中引理 3.3 的证明过程可得

$$\lim_{k\to\infty}\|\Delta_k\|=0 \tag{5}$$

从而

$$\lim_{k\to\infty}\|s_k\|=0 \tag{6}$$

于是由已知条件有

$$\lim_{k\to\infty}\|\nabla f(x_k)-g_k\|=0 \tag{7}$$

从而对充分大K，由(5)与条件(Ⅰ)和(Ⅲ)，有

$$-\phi_k(s_k)\geqslant c_1\epsilon_0\Delta_k\geqslant c_1\epsilon_0\|s_k\| \tag{8}$$

于是由(7)与条件(Ⅲ)，有

$$\left|\frac{f_k-f_{k+1}}{-\phi_k(s_k)}-1\right|=\left|\frac{f_k-f_{k+1}+\phi_k(s_k)}{-\phi_k(s_k)}\right|$$
$$\leqslant[\|\nabla f(x_k)-g_k\|+\|\nabla f(x_k)-\nabla f(x_k+\xi_k s_k)\|+\sigma_1\|s_k\|/2]/(c_1\epsilon_0)$$

其中$\xi_k\in(0,1)$．从而由(6)—(7)与$\nabla f(x)$一致连续可得

$$\lim_{k\to\infty}\frac{f_k-f_{k+1}}{-\phi_k(s_k)}=1 \tag{9}$$

由于对任意k，有$\rho_k=\dfrac{f_{l(k)}-f_{k+1}}{-\phi_k(s_k)}\geqslant\dfrac{f_k-f_{k+1}}{-\phi_k(s_k)}$．于是由(9)知，对充分大$k$有$\rho_k\geqslant\mu$．因此由算法 2 步 5 知，对充分大$k$有$\Delta_{k+1}\geqslant\Delta_k$，这与(5)产生了矛盾．即定理成立．

定理 2　设$\nabla f(x_k)$与g_k满足定理 3.1 中的条件，同时条件(Ⅰ)，(Ⅱ)，(Ⅲ)成立．那么

$$\lim_{k\to\infty}\|g_k\|=0$$

证　证明类似于文献[6]中的定理 3.2 证明．

定理 3　设$\{x_k\}$收敛于x^*，$\nabla^2 f(x^*)$是正定的，在(3)中$g_k=\nabla f(x_k)$，条件(Ⅰ)，(Ⅱ)，(Ⅲ)成立，$\nabla^2 f(x)$在x^*附近是 Lipschitz 连续的，并且如果当$\|s_k\|<\Delta_k$且$\lim\limits_{k\to\infty}\xi_k=0$时，有

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

与

$$\frac{\|B_k+g_k\|}{\|s_k\|}\leqslant\xi_k$$

则$\{x_k\}$超线性收敛于x^*．

证　证明类似于文献[6]中定理 3.7．

3　子问题的解

至于如何得到算法 2 中子问题(3)满足条件(Ⅰ)和(Ⅱ)的近似解s_k，我们直接用文献[6]中算法 4.1．具体

算法如下

算法3

步骤1　给出 $\xi_k > 0$，$\delta > 0$，$c > o$.

步骤2　令 $p_0 = 0$，$\gamma_0 = - g_k$，$d_0 = \gamma_0$，$i = 0$.

步骤3　计算 $\eta_i = d_i^T B_k d_i$. 如果 $\eta_i > \delta \| d_i \|^2$，转步骤4；如果 $0 < \eta_i \leqslant \delta \| d_i \|^2$，计算 $\alpha_i = \gamma_i^T \gamma_i / \eta_i$ 和 $p_{i+1} = p_i + \alpha_i d_i$. 如果 $\| p_{i+1} \| < \min\{\Delta_k, c \| g_k \|\}$，当 $i+1 < n$ 时，转步骤5，当 $i+1 = n$ 时，则停；否则，计算 $\tau > 0$ 使得 $\| p_i + \tau d_i \| = \min\{\Delta_k, c \| g_k \|\}$，令 $s_k = p_i + \tau d_i$ 并停止；如果 $\eta_i \leqslant 0$，计算 $\tau > 0$ 使得 $\| p_i + \tau d_i \| = \min\{\Delta_k, c \| g_k \|\}$，令 $s_k = p_i + \tau d_i$ 并停止.

步骤4　计算 $\alpha_i = \gamma_i^T \gamma_i / \eta_i$ 和 $p_{i+1} = p_i + \alpha_i d_i$. 如果 $\| p_{i+1} \| < \Delta_k$，当 $i+1 < n$ 时，转步骤5，当 $i+1 = n$ 时，则停；否则，计算 $\tau > 0$ 使得 $\| p_i + \tau d_i \| = \Delta_k$，令 $s_k = p_i + \tau d_i$ 并停止.

步骤5　计算 $\gamma_{i+1} = \gamma_i - \alpha_i B_k d_i$. 如果 $\| \gamma_{i+1} \| / \| g_k \| \leqslant \xi_k$，令 $s_k = p_{i+1}$，并停止；否则转步骤6.

步骤6　计算 $\beta_{i+1} = \gamma_{i+1}^T \gamma_{i+1} / \gamma_i^T \gamma_i$，$d_{i+1} = \gamma_{i+1} + \beta_i d_i$，令 $i = i+1$，$c = c + \delta^{-1}$，转步骤2.

4　数值试验

给出一些数值试验例子来比较 NTR 法与 UTR 法. 在这里假设

$$g_k = \nabla f(x_k),\ B_k = \nabla^2 f(x_k)$$

并令其中的参数为 $\mu=0.7$，$\eta=0.6$，$\epsilon=\delta=10^{-5}$，$\gamma_1=0.25$，$\gamma_2=0.75$，$\gamma_3=2$，$\xi_k=min\{\| g_k \|, 10^{-5}\}$，为方便起见，令向量的范数为 l_2 范数. 下面的数值试验是在相同终止准则下进行的比较. n_g 与 n_f 分别表示在迭代过程中梯度值和函数值计算的次数.

问题1　Rosenbrock 函数

$$f(x) = 100(x_2 - x_1{}^2)^2 + (1 - x_1)^2$$

$$x^* = (1, 1)$$

$$f(x^*) = 0$$

取 $\Delta_0 = 0.5$，$\overline{\Delta} = 1$，表1给出了对不同初始点，UTR 法($M = 0$) 与 NTR 法($M = 10$) 的测试结果.

表1　问题1的测试结果

初始点	$M = 0$(UTR)	$M = 10$(NTR)
$(-1.2, 1.0)^T$	$n_g = 24$，$n_f = 32$	$n_g = 15$，$n_f = 16$
$(2, -2)^T$	$n_g = 13$，$n_f = 15$	$n_g = 8$，$n_f = 8$
$(-3.635, 5.621)^T$	$n_g = 39$，$n_f = 50$	$n_g = 26$，$n_f = 27$
$(6.390, -0.221)^T$	$n_g = 19$，$n_f = 24$	$n_g = 13$，$n_f = 13$
$(1.489, -2.547)^T$	$n_g = 18$，$n_f = 23$	$n_g = 8$，$n_f = 8$

问题2　Cube 函数

$$f(x) = d(x_2 - x_1^3)^2 + (1 - x_1)^2$$

$$x^* = (1, 1)$$

$$f(x^*) = 0$$

取初始点 $x_0 = (-1.2, 1.0)^T$，$\Delta_0 = 1$，$\overline{\Delta} = 2$，表2给出了对不同 d 值 UTR 法($M = 0$) 与 NTR 法($M = 10$) 的测试结果.

表2　问题2的测试结果

d	$M = 0$(UTR)	$M = 10$(NTR)
10^2	$n_g = 28$，$n_f = 40$	$n_g = 8$，$n_f = 8$
10^4	$n_g = 110$，$n_f = 186$	$n_g = 8$，$n_f = 8$
10^6	$n_g = 583$，$n_f = 887$	$n_g = 6$，$n_f = 6$

从上述数值试验结果可以明显看出，NTR 法优于 UTR 法.

5　结　　论

本文给出解无约束最优化问题的非单调信赖域算法，并证明了它的收敛性. 有限的数值试验表明 NTR

法优于 UTR 法. 可以将这种非单调算法推广到解非光滑优化与变分不等式等问题.

参考文献：

[1] Dennis J E, Schnabel R B. Numerical Methods for Unconstrained Optimization and Nonlinear Equations[M]. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[2] Fletcher R. Practical Method of Optimization[M]. New York: 2nd Edition, 1987.

[3] Shultz G A, Schnabel R B, Byrd R H. A Family of Trust-Region-Based Algorithms for Unconstrained Minimization with Strong Global Convergence Properties[J]. SIAM Journal on Numerical Analysis, 1985, 22(1): 47－67.

[4] 袁亚湘, 孙文瑜. 最优化理论和方法[M]. 北京：科学出版社, 1997.

[5] Grippo L, Lampariello F, Lucidi S. A Nonmonotone Line Search Technique for Newton's Methods [J]. SIAM Journal on Numerical Analysis, 1986, 23(4): 707－716.

[6] Deng N Y, Xiao Y, Zhou F J. Nonmonotonic Trust Region Algorithm[J]. Journal of Optimization Theory and Applications, 1993, 76(2): 259－285.

[7] Grippo L, Sciandrone. Nonmonotone Globalization Techniques for the Barzilai-Borwein Gradient Method [J]. Computional Optimization and Applications, 2002, 23(2): 143－169.

[8] Zhou H C, Sun W. Nonmonotone Descent Algorithm for Nonsmooth Unconstrained Optimization Problems[J]. International Journal of Pure and Applied Mathematics, 2003, 9(2): 153－163.

[9] Sun W. Non-Monotone Trust Region Method for Optimization [J]. Applied Mathematics and Computation, 2004, 156(1): 159－174.

[10] Zhou Q Y, Sun W. An Adaptive Nonmonotone Trust Region Method with Curvilinear Searches [J]. J Comput Math, 2006, 24(6): 761－770.

[11] Chamberlain R M, Powell M J D, et al. The Watchdog Technique for Forcing Convergence in Algorithms for Constrained Optimization [J]. Mathematical Programming Study, 1982, 16: 1－17.

[12] Grippo L, Lampariello F, Lucidi S. A Truncated Newton Method with Nonmonotone Line Search for Unconstrained Optimization [J]. Journal of Optimization Theory and Application, 1989, 60(13): 401－419.

Nonmonotonic Trust Region Method for Unconstrained Optimization

YANG Zheng-hao

School of Mathematics and Computer Science, Nanjing Normal University, Nanjing, 210097

Abstract: In order to guarantee the global convergence of an algorithm for unconstrained optimization, the usual trust region methods force a monotonical decrease of the objective function at each step, which sometimes can considerably slow the rate of convergence. A nonmonotonic trust region method presented in this paper allows an increase in the function value at some step, while retaining global convergence. Finally, numerical experiments show that the nonmonotonic trust region method is superior to the usual trust region methods.

Key words: unconstrained optimization; trust region method; nonmonotone method; convergence

责任编辑　张　枸