三项预处理共轭梯度法与信赖域子问题

后六生,孙文瑜

(南京师范大学数学与计算机科学学院,南京,210097)

[摘要]　信赖域方法是解无约束优化问题的有效的和可靠的方法.共轭梯度法由于不需要矩阵计算和存贮,成了解大型问题的首选方法.在本文中,我们提出了解信赖域子问题的三项预处理共轭梯度法,并将这个方法嵌入解大型最优化问题的信赖域算法中.文章讨论了方法的特性,证明了方法的总体收敛性质,并给出了有限的数值试验.

[关键词]　三项共轭梯度法;信赖域方法;算法

[中图分类号]O224;　[文献标识码]A;　[文章编号]1001-4616(2001)03-0001-06

0　引言

考虑求解无约束最优化问题

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

这里 $f(x)$ 是连续可微函数.

信赖域方法是目前解无约束最优化问题(1) 的最重要的方法之一.这个方法具有很强的总体收敛性质和较快的局部收敛速度.不像线搜索方法每次仅在一维情形获得模型问题的一个极小点,信赖域方法每次迭代总是寻找一个以 x_k 为中心的广义球

$$\{x_k + d \mid \|d\| \leqslant \Delta_x\}$$

上的"最好的点".假定在这个广义球内,二次模型是目标函数的一个合适的模拟,我们就利用 n 维二次模型的极小点 s_k 作为搜索方向;否则,我们就拒绝利用这个方向.

在信赖域方法中,如何求解信赖域子问题是一个关键问题.通常采用的信赖域子问题是基于二次模型的,即

$$\begin{aligned}\min\quad & q_k(s) = g_k^T s + \frac{1}{2} s^T B_k s \\ s.t.\quad & \|s\| \leqslant \Delta_k,\end{aligned} \tag{2}$$

其中,g_k 是目标函数在当前点的梯度,B_k 是目标函数的 Hesse 近似,$\Delta_k > 0$ 是信赖域半径.目前,解子问题(2) 的主要方法是牛顿法和拟牛顿法.

然而,当变量的个数变得很大时,牛顿法和拟牛顿法的效果就受到了影响.在这种情形,共轭梯度法就成了处理信赖域子问题的理想方法.

Steihaug(1983) 提出了用预处理共轭梯度法(preconditioned CG) 解信赖域子问题.在这个著名的工作中,他设计了三个停止准则,即除了正常情况下,按照预处理CG方法得到搜索方向

收稿日期:2001-02-16
作者简介:后六生,1973—,南京师范大学数学与计算机科学学院硕士研究生,从事计算数学的学习与研究.

(或搜索点) s_k 外,当搜索方向 s_k 在信赖域外或 s_k 是模型 Hesse 的负曲率方向时,都要求新的解点直接在信赖域边界上,即求 $\alpha > 0$ 使得

$$\| s_k + \alpha d_k \| = \Delta_k,$$

并令 $s_{k+1} = s_k + \alpha d_k$ 作为下一次子问题的搜索方向. 类似于通常的信赖域方法,对于成功迭代,Steihaug 方法接受子问题的解作为下一次迭代方向. 对于不成功迭代,$x_{k+1} = x_k$,丢掉了 s_k 所含有的有用信息. 对于某些问题,如果频繁地不满足下降性条件,频繁地令 $x_{k+1} = x_k$,便大大减弱了方法的效率,使方法难以较快前进. 在本文中,我们企图通过采用三项预处理 CG 方法来克服这个问题.

对于解信赖域子问题,充分下降性条件是重要的. 设目标函数 $f(x)$ 的实际下降为

$$\mathrm{Ared}(s_k) = f(x_k) - f(x_k + s_k) \tag{3}$$

目标函数 $f(x)$ 的预测下降,即模型函数 $q_k(s)$ 的下降为

$$\mathrm{Pred}(s_k) = q_k(0) - q_k(s_k) = - q_k(s_k) \tag{4}$$

Powell(1975) 证明,如果信赖域步长 s_k 满足

$$\mathrm{Pred}(s_k) \geqslant c \| g_k \| \min \{\Delta_k, \| g_k \| / \| B_k \|\} \tag{5}$$

则求解最优化问题 (1) 的信赖域算法产生的点列收敛到目标函数的极小点. (5) 给出了每步迭代中模型函数下降的一个下界. 从 (5) 和信赖域算法,我们立即得到

$$\mathrm{Ared}(s_k) = f(x_k) - f(x_k + s_k) \geqslant c\mu \| g_k \| \min \{\Delta_k, \| g_k \| / \| B_k \|\} \tag{6}$$

(6) 给出了信赖域算法在每步迭代中目标函数的实际下降的一个下界. 这个实际下降表达式 (6) 对于我们证明信赖域方法的总体收敛性是至关重要的.

本文组织如下:本节引言之后,在第一节描述了三项预处理共轭梯度法解信赖域子问题的算法,并给出了收敛性证明;在第二节,我们给出初步的数值试验结果. 结果表明,三项预处理共轭梯度法有一定的优越性.

1　算法和收敛性

调比或预处理技术对于最优化算法是十分重要的,同样地,它们在信赖域方法中也是十分重要的. 例如,当模型 Hesse 是坏条件时,即其特征值分布很分散,采用调比技术,或预处理技术,或加权范数,可以有效地改善其特征值结构,进而改善算法的收敛性质. 为此,我们采用三项预处理共轭梯度法解信赖域子问题.

考虑信赖域子问题

$$\begin{aligned} \min \quad & q(s) = g^T s + \frac{1}{2} s^T B s, \\ s.t. \quad & \| s \|_C \leqslant \Delta \end{aligned} \tag{7}$$

其中 $\| s \|_C^2 = s^T C s$,是向量的 C - 范数,C 是 $n \times n$ 对称正定矩阵. 为方便起见,在(7) 中我们暂时忽略下标.

Beale(1972) 提出了三项共轭梯度法,我们对其进行预处理加工. 并将它融入解信赖域子问题中. 这样,我们得到的下述算法就是 Steihaug 算法的一个改进.

算法 1(解信赖域子问题).

步 0　给出 $g \in R^n$,$B \in R^{n \times n}$,$\varepsilon > 0$,RE 为布尔控制变量,如果 $RE = .true.$,则 $s_1 = 0$,否则 $s_1 = s^*$. $g_1 = g$,$v_1 = C^{-1} g_1$,$d_1 = - v_1$,$k = 1$.

步 1　如果 $\|g_k\| \leqslant \varepsilon$,令 $s^* = s_k$,停止;否则,计算 $d_k^T B d_k$;如果 $d_k^T B d_k \leqslant 0$,转步 3;

计算 $\alpha_k = g_k^T v_k / d_k^T B d_k$.

步 2　如果 $\|s_k + \alpha_k d_k\| \geqslant \Delta$. 转步 3;否则,$s_{k+1} = s_k + \alpha_k d_k$,$g_{k+1} = g_k + \alpha_k B d_k$,$v_{k+1} = C^{-1} g_{k+1}$,$\beta_k = g_{k+1}^T v_{k+1} / g_k^T v_k$,

$$\gamma_k = \begin{cases} 0, & \text{当 } k = t+1 \text{ 时} \\ \dfrac{g_{k+1}^T C^{-1}(g_{t+1} - g_t)}{d_t^T C^{-1}(g_{t+1} - g_t)}, & \text{当 } k > t+1 \text{ 时}. \end{cases}$$

$d_{k+1} = -v_{k+1} + \beta_k d_k + \gamma_k d_t$,$k = k+1$,转步 1.

步 3　解 $\|s_k + \alpha d_k\| = \Delta$ 得到 $\alpha_k \geqslant 0$. 令 $s^* = s_k + \alpha_k d_k$. 停止.

在以上算法中,如果 d_k 是负曲率方向,或者所产生的 $s_k + \alpha_k d_k$ 在信赖域外面,我们都令新点 $s^* = s_k + \alpha k d_k$ 在信赖域边界上. 将上述算法融入信赖域算法,我们就得到极小化问题(1)的采用三项预处理 CG 方法的信赖域算法.

算法 2(采用三项预处理 CG 方法的信赖域算法).

步 0　初始点:给出初始点 x_1,初始信赖域半径 $\Delta_1 > 0$. $\varepsilon \geqslant 0$,$0 < \mu < \eta < 1$,$0 < \gamma_1 < 1 < \gamma_2$. 令 $k = 0$.

步 1　如果 $\|\nabla f(x_k)\| \leqslant \varepsilon$,停止.

步 2　用三项预处理 CG 算法 1 解 TR 子问题,得到近似解 s_k.

步 3　计算

$$r_k = \frac{f(x_k) - f(x_k + s_k)}{-q_k(x_k + s_k)}.$$

令

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{如果 } r_k > \mu \\ x_k, & \text{否则}. \end{cases}$$

步 4　校正信赖域半径:

$$\Delta_{k+1} = \begin{cases} \gamma_1 \Delta_k, & \text{如果 } r_k > \mu, \\ \Delta_k, & \text{如果 } r_k \in [\mu, \eta), \\ \gamma_2 \Delta_k, & \text{如果 } r_k \geqslant \eta. \end{cases}$$

步 5　校正 B_k;令 $k = k+1$,转步 1.

显然,在不成功情形,$x_{k+1} = x_k$. 如果这种情形频繁地发生,算法的效率将大受影响. 为了对付这种情形,我们可以在算法中采取下面两种措施:

措施一:减小 μ,例如取 $\mu = 0.01$ 或 0.001. 在我们的算法中,$\mu = 0.01$,$\eta = 0.9$,$\gamma_1 = \dfrac{1}{2}$,$\gamma_2 = 2$.

措施二:我们认为,在不成功情形,$x_{k+1} = x_k$,将丢掉计算得到的 s^* 所含有的有用信息. 为此,我们可以在不成功迭代时,以 s^* 为再开始方向,用三项共轭梯度法再做,期望得到较好的效果. 我们的算法中,设置了布尔控制变量 *RE*,在不成功迭代 *RE* = . *FAULSE*. ,这时以 s^* 为再开始方向进入三项 CG 算法子程序;否则,*RE* = . *TRUE*. 在下一轮解子问题算法中,$s^* = 0$.

我们首先证明我们的算法满足充分下降性条件,然后给出总体收敛性定理.

定理 1　设 s^* 是(7)的近似解,则

$$\mathrm{Pred}_k = -q_k(s^*) \geqslant \frac{1}{2}c_0\|\hat{g}\|\min\{\Delta,\frac{\|\hat{g}\|}{\|\hat{B}\|}\} \tag{8}$$

这里 $\hat{g} = D^{-T}g$,$\hat{B} = D^{-T}BD^{-1}$,$C = D^TD$.

证明　显然子问题(7)

$$\min \quad g^Ts + \frac{1}{2}s^TB_s$$
$$s.t. \quad \|s\|_C \leqslant \Delta$$

等价于

$$\min \quad g^Ts + \frac{1}{2}s^TB_s$$
$$s.t. \quad \|Ds\| \leqslant \Delta \tag{9}$$

令 $w = Ds$,$\hat{g} = D^{-T}g$,$\hat{B} = D^{-T}BD^{-1}$,则(8) 等价于

$$\min \quad \hat{q}_k(w) = \hat{g}^Tw + \frac{1}{2}w^T\hat{B}w$$
$$s.t. \quad \|w\| \leqslant \Delta \tag{10}$$

这样,我们容易得到下面的不等式

$$-q_k(s^*) = -\hat{q}_k(w^*) \geqslant c_0[\hat{q}_k(0) - \hat{q}_k(\bar{w})] \geqslant c_0[\hat{q}_k(0) - \hat{q}_k(-\alpha\frac{\Delta}{\|\hat{g}\|}\hat{g})]$$
$$\geqslant c_0[\|\hat{g}\|\Delta - \frac{1}{2}\alpha^2\|\hat{B}\|\Delta^2] \geqslant \frac{1}{2}c_0\|\hat{g}\|\min\{\Delta,\frac{\|\hat{g}\|}{\|\hat{B}\|}\}.$$

其中,w^* 是 $\hat{q}_k(w)$ 的近似解,$\bar{\omega}$ 是 $\hat{q}_k(w)$ 的精确解.

注意到用三项共轭梯度法解子问题产生的序列 $\{s^*\}$ 满足近似解的要求,因此算法 1 产生的 s^* 满足定理 1 中的不等式(8).

下面,我们给出算法的总体收敛性定理. 我们假定

$$\|\hat{B}\| \leqslant \sigma, \|D^{-1}\| \leqslant \sigma_1$$

其中 σ 和 σ_1 是正常数.

定理 2　如果 $f: R^n \to R$ 连续可微且下有界,那么由算法 2 产生的序列满足

$$\lim_{k\to+\infty}\inf\|\nabla f(x_k)\|_{C_k^-} = 0 \tag{11}$$

这里 $C_k = D_k^TD_k$.

证明　要证

$$\lim_{k\to+\infty}\inf\|\nabla f(x_k)\|_{C_k^{-1}} = 0$$

即证

$$\lim_{k\to+\infty}\inf\|\hat{g}(x_k)\| = 0 \tag{12}$$

今假定存在 $\varepsilon > 0$,使得对于充分大的 k,$\|\hat{g}(x_k)\| \geqslant \varepsilon$. 现在我们来证明

$$\sum_{k=1}^{+\infty}\Delta_k < +\infty \tag{13}$$

成立.

如果算法只有有限多次迭代是成功迭代,那么对于充分大的 k,$\Delta_{k+1} \leqslant \gamma_1\Delta_k$,从而(13) 显然成立. 如果存在无限多次迭代成功迭代,由于我们的算法满足定理 1 的结论,故从(8) 可知,无穷

序列 $\{k_i\}$ 满足

$$\sum_{i=1}^{+\infty}\Delta_{k_i} < +\infty$$

由算法 2 和 Powell(1975),得到

$$\sum_{k=1}^{+\infty}\Delta_k \leqslant \left(1+\frac{\gamma_2}{1-\gamma_1}\right)\sum_{i=1}^{+\infty}\Delta_{k_i} \tag{14}$$

从而(13) 成立.

此外,从我们的算法 1 可知,无论在哪种情形终止,我们都有

$$\| x_{k+1}-x_k\| \leqslant \Delta_k \tag{15}$$

从而由$\Delta_k \to 0$ 得到 $\{x_k\}$ 收敛.进一步,若$\|\hat{B}\| \leqslant \sigma$,$\|D^{-1}\| \leqslant \sigma_1$ 由(10) 有

$$|\hat{q}(w^*)-\hat{g}^T w^*| \leqslant \frac{1}{2}\sigma\|w^*\|^2 \leqslant \frac{1}{2}\sigma\Delta_k^2.$$

于是

$$|q_k(s^*)-g^T s^*| \leqslant \frac{1}{2}\sigma\Delta_k^2 \tag{16}$$

从而存在 $\varepsilon_k \to 0$,满足

$$|f(x_k+s^*)-f(x_k)-q_k(s^*)| \leqslant \varepsilon_k\Delta_k \tag{17}$$

但是,定理 1 告诉我们

$$-q_k(s^*) \geqslant \frac{1}{2}c_0\tau\Delta_k,\text{对某个 } \tau > 0.$$

从而,(17) 两边同除以 $-q_k(s^*)$,并令 $k\to\infty$,得

$$r_k \to 1.$$

这表明 $\Delta_{k+1} \geqslant \Delta_k$,即 $\{\Delta_k\}$ 不可能收敛到零.这与(13) 矛盾.这个矛盾证明了定理的结论正确.

2 数值结果

我们对 Steihaug 算法和改进的算法分别编程进行了有限的数值试验,一个函数是优化的标准测试函数 Beale function,另一个函数为 $f(x)=(x_1-2)^2+(x_2-10)^4$,初始点为(1,1).实验结果如表 1.

表 1 数值试验结果

算法	迭代次数	终值点	梯度值
Beale function			
Steihaug 算法	146	(3.00039,0.50010)	(-1.25210 e-5,1.11417 e-4)
改进的算法	72	(2.99993,0.49998)	(-1.34187 e-10,-9.72349 e-6)
$f(x)=(x_1-2)^2+(x_2-10)^4$			
算法	迭代次数	终值点	梯度值
Steihaug 算法	222	(1.99999,9.97103)	(-1.15145 e-8,1.81393 e-3)
改进的算法	19	(2.00000,9.98715)	(4.21884 e-13,-8.47405 e-6)

从有限的数值例子可以看出改进的算法比 Steihaug 算法较好.进一步的数值试验仍需进行.

[参考文献]

[1] Beale E M L. A derivation of conjugate gradients. in :F. A. Lootsma eds., Numerical Methods for Nonlinear Optimization[M]. Academic Press :London, 1972. 39—43.

[2] Gay D M. Computing optimal locally constrained steps[J]. SIAM J. Sci. Stat. Comp., 1981, 2 :186—192.

[3] Gould N I M, Lucidi S, Roma M. Toint P L. Solving the trust region subproblem using the Lanczos method[J]. SIAM J. Optimization, 1999, 9(2) :504—525.

[4] More J J. Recent Developments in Algorithms and Software for Trust Region Methods, in A. Bachen, etc. eds., Mathmatical Programming[M]. The State of the Art, Springer. Verlag, Berlin, 1983.

[5] More J J, Soresen D C. Computing a trust region step[J]. SIAM J. Sci. Stat. Comp., 1983, 2 :186—197.

[6] Powell M J D. A new algorithm for unconstrained optimization. in Nonlinear Programming. J. B. Rosen, O. L. Mangnsatian ang K. Ritter, eds[M]., Academic Press :New York, 1970. 31—65.

[7] Steihaug T. The conjugate gradient method and trust regions in lager scale optimization[J]. SIAM J. Numer. Anal. 1983, 20 :626—637.

[8] Ph L. Toint. Towards an efficient sparcity exploiting Newton method for minimization. In :I. Duff. ed., Sparse Matrices and Their Uses[M]. Academic Press, 1981. 57—88.

[9] 袁亚湘,孙文瑜. 最优化理论与方法[M]. 北京 科学出版社,1997.

Three-Term Preconditioned Conjugate Gradient Method and Trust Region Subproblem

Hou Liusheng, Sun Wenyu

(School of Mathematics and Computer Science, Nanjing Normal University, Nanjing, 210097, PRC)

Abstract :Algorithms based on trust regions have been shown to be robust for unconstrained optimization problems. Steihaug has shown that an approximate solution of the trust region problem may be found by the PCG method. The aim of this paper is to construct a three-term conjugate gradient method to solve the trust region subproblem. Our algrithm is an improvement of Steihaug's algorithm.

Key words :three-term preconditioned conjugate gradient method ;trust region method ;optimization

[责任编辑:陆炳新]