DOI: 10.13656/j.cnki.gxkx.2009.04.002
广西科学 Guangxi Sciences 2009, 16(4): 397～399

一个修改的 BFGS信赖域算法*

A Modified BFGS-Trust-Region Algorithm

袁功林,吴燕林,韦增欣
YUAN Gong-lin, WU Yan-lin, WEI Zeng-xin

(广西大学数学与信息科学学院,广西南宁 530004)
(College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, 530004, China)

摘要:给出一个修改的 BFGS校正信赖域算法,并分析其收敛性.该算法能够保持校正矩阵正定和收敛速度是二次的.

关键词: BFGS方法 信赖域方法 全局收敛性 无约束优化
中图法分类号: O242.23 文献标识码: A 文章编号: 1005-9164(2009)04-0397-03

Abstract A modified BFGS-trust-region algorithm is proposed and the convergence is studied. This algorithm ensure that the update matix is positive definite and the convergence rate is quadratic.

Key words BFGS method, Trust-region method, global convergence, unconstrained optimization

对于无约束最优化问题:

$$\min f(x) \quad x\in R^n, \tag{0.1}$$

其中 $f(x)$ 是连续可微函数,目前 BFGS方法是拟牛顿方法中解问题(0.1)的最有效方法.文献[1～3]给出一些修改的 BFGS方法并分析其收敛性.针对著名的 BFGS校正公式

$$B_{k+1} = B_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \tag{0.2}$$

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,Li和 Fukashima[3]给出一个修改的 BFGS校正公式:

$$B_{k+1} = B_k + \frac{y_k^* {y_k^*}^T}{s_k^T y_k^*} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \tag{0.3}$$

其中 $y_k^* = y_k + t_k\|\nabla f(x_k)\| s_k$,$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$,$t_k = C + \max\{0, -\frac{y_k^T s_k}{\|s_k\|^2}\}\|\nabla f(x_k)\|^{-1}$,$C > 0$ 是一个常数,并且分析其收敛性.我们知道信赖域算法是非常重要的算法,关于这类算法的研究见文献[4～8]等.本文将公式(0.3)与一般的信赖域算法相结合,给出一个新的算法,此算法能够保持校正矩阵的正定性,并且保证收敛速度是二次的.我们记 $\|\cdot\|$ 是 R^1 中的 Euclid范数,$g(x)\in R^n$ 是 f 在 x 处的梯度值,$\{x_n\}$ 是由算法产生的点列,记 $f_k := f(x_k)$,$g_k := \nabla f(x_k)$.

1 算法及其性质

给出解问题(0.1)的一般信赖域算法子问题的定义式为

$$\min Q_k(s) = f_k + g_k^T s + \frac{1}{2} s^T B_k s \quad s\in R^n \quad \text{s.t.} \|s\| \leqslant \Delta_k, \tag{1.1}$$

其中,B_k 是 $f(x)$ 在 x_k 的 Hesse矩阵或其近似,$\Delta_k > 0$ 是信赖域半径,s 是试探步.求解信赖域子问题是非常重要的,文献[9, 10]给出了问题(1.1)相应的方法并讨论其收敛性.

设 $\triangle f_k$ 是 f 在第 k 步的实际下降量,定义

$$\triangle f_k = f_k - f(x_k + s_k), \tag{1.2}$$

$\triangle j_k$ 为对应的预测下降量:

$$\triangle j_k = f_k - Q_k(s), \tag{1.3}$$

定义比值

$$r_k = \frac{\triangle f_k}{\triangle j_k}, \tag{1.4}$$

利用文献[11]中所给的信赖域算法与(0.3)式结合,

收稿日期: 2008-09-15
修回日期: 2009-02-13
作者简介: 袁功林(1976-),男,副教授,主要从事优化理论与方法的研究.
* This work is supported by China NSF grands 10761001 and the Scientific Research Foundation of Guangxi University (Grant No. X081082).

给出一个修改的 BFGS校正的信赖域算法.

算法 1

步骤 0　给定 $x_0 \in R^n$,对称阵 $B_0 \in R^{n \times n}$,$\triangle_0 > 0$;

步骤 1　计算 g_k.如果满足终止规则,则终止;否则转步骤 2;

步骤 2　解子问题(1.1),得试探步 d_k;

步骤 3　计算 $f(x_k + s_k)$ 和 r_k 的值;

步骤 4　如果 $r_k < 0.25$,则令 $\triangle_{k+1} = \| s_k \| /4$;如果 $r_k > 0.75$和 $\| s_k \| = \triangle_k$,令 $\triangle_{k+1} = 2\triangle_k$;否则置 $\triangle_{k+1} = \triangle_k$;

步骤 5　若 $r_k \leqslant 0$,置 $x_{k+1} = x_k$;否则置 $x_{k+1} = x_k + s_k$,利用校正公式(0.3)产生 B_{k+1},令 $k := k+1$,转步骤 1.

由算法 1可得到性质:

注 1　用 BFGS校正方法,当 $y_k^T s_k > 0$时,B_k保持正定性.当 B_k正定时,有

$$y_k^{*T} s_k = (y_k + t_k \| \nabla f(x_k) \| s_k)^T s_k \geqslant C \| \nabla f(x_k) \| \| s_k \|^2 > 0, \qquad (1.5)$$

其中 $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$,$t_k = C + \max\{0, -\frac{y_k^T s_k}{\| s_k \|^2}\} \| \nabla f(x_k) \|^{-1}$,$C > 0$是一个常数,所以 B_{k+1}也是正定的.当 B_0取对称正定矩阵时,用数学归纳法可以证明由校正公式(0.3)产生的 B_k都是正定的.相应地,若 $\lim_{k \to \infty} x_k = x^*$,得到 $\nabla^2 f(x^*)$是正定的,$B^* = \nabla^2 f(x^*)$.

2　算法的收敛性分析

此节讨论的 B_k都是在 B_0为对称正定时产生的.首先,定义水平集

$$K = \{x \in R^n: f(x) \leqslant f(x_0)\}.$$

为了证明算法 1的收敛性,需要假设条件:

(A) f在 K上二次连续可微,且存在 $M > 0$使得 $\| B_k \| < M$成立.

注 2　根据注 1的讨论知 B_k是正定的,再由假设(A)可得,存在 $m > 0$使得

$$t^T \nabla^2 f(x) t \geqslant m \| t \|^2, \forall x \in K, t \in R^n,$$

所以根据文献[11]中的定理 1.3.16可得到水平集 K是有界闭凸集.

注 3　利用 K是有界集与 $f(x_k)$是单调下降的,可知由算法 1产生的点列 $\{x_k\} \subset K$,从而 $\{x_k\}$是有界点列,所以 $\{x_k\}$存在聚点,设 $x_k \to x^*$.

定理 1　根据上述讨论和假设(A),对于主序列有 $r_k \to 1$,$x_k \to x^*$,$glb(x_k) > 0$(glb表示 x_k的总体最小界),以及对于充分大的 k,约束 $\| d \|_2 < \triangle_k$,那么收敛速度是二阶的.

证明　由算法 1产生的序列中存在一个子序列,满足下列情况之一:

(1) $r_k < 0.25$,$\triangle_{k+1} \to 0$(因而 $\| s_k \| \to 0$),

(2) $r_k \geqslant 0.25$,$glb(\triangle_k) > 0$.

假定情况(1)中的子序列 $x_k \to x^*$.考虑方向 $d_k = -B_k^{-1} g_k$.由于 B^*正定,则对充分大的 k,方向 d_k为下降方向.若 $\| d_k \| \leqslant \triangle_k$,则 $s_k = d_k$,

$$\triangle j_k = f_k - Q(s_k) = -g_k^T s_k - \frac{1}{2} s_k^T B_k s_k = (B_k s_k)^T s_k - \frac{1}{2} s_k B_k s_k = \frac{1}{2} s_k^T B_k s_k \geqslant \frac{1}{2} \lambda_k \| s_k \|^2, \quad (2.1)$$

其中 λ_k是 B_k的最小特征值.若 $\| d_k \| \geqslant \triangle_k$,则由 j_k的最优性,

$$\triangle j_k \geqslant f_k - Q(\triangle_k d_k / \| d_k \|) \geqslant -\frac{\triangle_k d_k^T g_k}{\| d_k \|} - \frac{1}{2} \frac{\triangle_k^2 d_k^T B_k d_k}{\| d_k \|^2} = \frac{\triangle_k d_k^T B_k d_k}{\| d_k \|} - \frac{1}{2} \frac{\triangle_k^2 d_k^T B_k d_k}{\| d_k \|^2} \geqslant \frac{1}{2} \frac{\triangle_k d_k^T B_k d_k}{\| d_k \|} \geqslant \frac{1}{2} \triangle_k \lambda_k \| d_k \| \geqslant \frac{1}{2} \triangle_k^2 \lambda_k = \frac{1}{2} \lambda_k \| s_k \|^2. \qquad (2.2)$$

(2.1)式和(2.2)式表明 $\triangle_k \geqslant \frac{1}{2} \lambda_k \| s_k \|^2$总成立.注意到 $\triangle f_k = \triangle j_k + o(\| s_k \|^2)$,从而得到

$$r_k = \frac{\triangle f_k}{\triangle j_k} = 1 + \frac{o(\| s_k \|^2)}{\triangle j_k} \to 1, \qquad (2.3)$$

这与 $r_k < 0.25$矛盾.所以序列 $x_k \to x^*$不属于情况(1),而属于情况(2).于是,如果 k充分大使得 $\| x_k - x^* \| \leqslant \frac{1}{2} glb(\triangle_k)$,则文献[11]中的牛顿收敛性定理 3.2.1成立,故 $x_{k+1} = x_k - B_k^{-1} g_k$满足 $\| x_{k+1} - x^* \| < \| x_k - x^* \|$,并且在 $\Omega_k = \{x | \| x - x_k \| \leqslant \triangle_k\}$中可行.因此整个序列 $x_k \to x^*$.由(2.3)可知整个序列满足 $r_k \to 1$,且在情况(2)下有 $glb(\triangle_k) > 0$.由于 k充分大时,$\| s_k \| < \triangle_k$,算法 1化为牛顿迭代,所以由文献[11]中的定理 3.2.1可知序列 $\{x_k\}$具有二次收敛速度.

参考文献:

[1] Powell M J D. A new algorithm for unconstrained optimation[M]//Rosen J B, Mangasarian O L, Ritter K, et al. Nonlinear Programming. New York: Academic Press, 1970.

[2] Wei Z, Yu G, Yuan G, et al. The superlinear convergence of a modified BFGS-type method for unconstrained optimization[J]. Computational Optimization and Applications, 2004, 29: 315-332.

[3] D LI, Fukushima M. A modified BFGS method and its

global convergence in nonconvex minimization [J]. Journal of Computational and Applied Mathematics, 2001, 129: 15-35.

[4] Byrd R H, Schnabel R B, Shultz G A. Approximate solution of the trust region problem by minimization over two-dimensional subspaces[J]. math Prog, 1988, 40: 247-263.

[5] Buleau J P, Vial J Ph. Curvilinear path and trust region in unconstrained optimization, a convergence analysis [J]. Math Prog Study, 1987, 30: 82-101.

[6] 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, 200.

[7] Shultz G A, Schnabel R B, Byrd R H. A family of trust-region-based algorithms for unconstrained miimization with strong global convergence properties[J]. SIAM J Numer Anal, 1985, 22: 47-67.

[8] 袁亚湘.信赖域方法的收敛性 [J].计算数学, 1994, 16: 333-346.

[9] Nocedal J, Yuan Y X. Combining trust-region and line search techniques [J]. Advance in Nonlinear Programming, Applied Optimization, Kluwer Academic Publ, Dordrecht, 1998, 14: 153-175.

[10] Rendl F, Wolkowicz H. A semidefinite framework for trust region subproblems with applications to large scale minimization[J]. Math Program, 1997, 77: 273-299.

[11] Yuan Y, Sun W. Theory and methods of optimization [M]. Beijing: Science Press of China, 1999.

(责任编辑: 尹 闯)

(上接第 396页 Continue from page 396)

参考文献:

[1] Stynes M, E O'Riordan. A finite element method for a singularly perturbed boundary value problem [J]. Numer Math, 1986, 50: 1-15.

[2] Stynes M, E O'Riordan. An analysis of a singularly perturbed two-point boundary value problem using only finite element techniques[J]. Math Comp, 1991, 56: 663-675.

[3] Sun G, Stynes M. Finite element methods for singularly perturbed high-order elliptic two-point boundary problems, I: Reaction-diffusion-type problems[J]. IMA J Numer Anal, 1995, 15: 117-139.

[4] Sun G, Stynes M. Finite element methods for singularly perturbed high-order elliptic two-point boundary problems, II: Convection-diffusion-type problems [J]. IMA J Numer Anal, 1995, 15: 197-219.

[5] Liu W, Tang T. Error analysis for a Galerkin spectral method with coordinate transformation for solving singularly perturbed problems [J]. Appl Numer Math, 2001, 38: 315-345.

[6] H J J Miller, E O'Riordan, G I Shishkin. Fitted numerical methods for singularly perturbation problems [M]. Singapore: World Scientific, 1996.

[7] Tang T, Trummer M R. Boundary layer resolving pseudospectral methods for singular perturbation problems[J]. SIAM J Sci Comput, 1996, 17: 430-438.

[8] Qiu Y, Sloan D M. Analysis of difference approximations to a singular perturbed two-point boundary value problem on an adaptively generated grid [J]. J Comput Appl Math, 1999, 101: 1-25.

[9] Roos H G, Stynes M, Tobiska L. Numerical methods for singularly perturbed differential equations [M]. Berlin: Springer-Verlag, 1996.

[10] Shishkin G I. Grid approximation of singularly perturbed parabolic equations with internal layers[J]. Sov J Numer Anal Math Model, 1988, 3: 393-407.

[11] Liu S, Xu Y. Galerkin methods based on Hermite splines for singular perturbation problems[J]. SIAM J Number Anal, 2006, 43: 2607-2623.

[12] R E O'Malley. Introduction to singular perturbations [M]. New York: Academic Press, 1974.

[13] Li R, Chen Z, Wu W. Generalized difference methods for differential equations: numerical analysis of finite volume methods [M]. New York: Marcel Dekker, Inc, Basel, 2000.

(责任编辑: 尹 闯)