DOI: 10.13656/j.cnki.gxkx.2004.03.010
广西科学 Guangxi Sciences 2004, 11 (3): 195~196, 200

一个新的 BFGS信赖域算法*
A New BFGS Trust Region Algorithm

袁功林　　韦增欣
Yuan Gonglin　Wei Zengxin

(广西大学数学与信息科学学院　南宁市大学路 100号　530004)
(Coll. of Math., & Info. Sci., Guangxi Univ., 100 Daxuelu, Nanning, Guangxi, 530004, China)

摘要　给出能够保持校正矩阵是正定的新的 BFGS信赖域算法，以及该算法的全局收敛性和其二次收敛速度.
关键词　信赖域算法　BFGS方法　全局收敛性　收敛速度
中图法分类号　O241.7
Abstract　A new BFGS-trust-region algorithm which ensures a positive update matrix was given. It's global convergence and quadratieally convergent speed were given.
Key words　trust-region algorithm, BFGS method, global convergence, convergence speed

1　BFGS校正公式

对于无约束最优化问题

$$\min f(x), x \in R^n, \tag{1.1}$$

其中 $f(x)$ 是连续可微函数.信赖域算法是非常重要的算法,文献 [1~4]分析这类算法的性质.众所周知 BFGS方法是拟牛顿方法中解无约束最优化问题 (1.1) 的重要的方法[5~7],文献 [8, 9]分析拟牛顿方法收敛性. BFGS校正公式为

$$B_{k+1} = B_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \tag{1.2}$$

其中 B_k 是 f 在 x_k 处的 Hesse矩阵或其近似矩阵; $s_k = x_{k+1} - x_k$; $y_k = g_{k+1} - g_k$, g_k 和 g_{k+1} 分别是 $f(x)$ 在 x_k 和 x_{k+1} 处的梯度值. 以 y_k^* 代替 y_k, 即 $y_k^* = \frac{y_k^T s_k}{|y_k^T s_k|} y_k$,相应地可得到一个新的 BFGS校正公式

$$B_{k+1} = B_k + \frac{y_k^* y_k^{*T}}{s_k^T y_k^*} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \tag{1.3}$$

其中 $y_k^* = \frac{y_k^T s_k}{|y_k^T s_k|} y_k$. 本文将公式 (1.3)与一般的信赖域算法相结合,得到一个新的 BFGS信赖域算法,新的算法能够保证校正矩阵是正定的.为叙述方便,本文采用下列记号: $\|\cdot\|$ 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 $f_k := f(x_k)$, $g_k := \nabla f(x_k)$, $B_k = \nabla^2 f(x_k)$.

2　新的 BFGS信赖域算法

问题 (1.1) 的一般信赖域算法子问题的定义式为

$$\min p_k(s) = f_k + g_k^T s + \frac{1}{2} s^T B_k s,$$
$$\text{s.t.} \|s\| \leqslant \Delta_k, \tag{2.1}$$

其中 $s \in R^n$ 是试探步; $\Delta_k > 0$是信赖域半径.设 Δf_k 是 f 在第 k步的实际下降量,定义为: $\Delta f_k = f_k - f(x_k + s_k)$, Δp_k 为对应的预测下降量: $\Delta p_k = f_k - p_k(s)$,定义比值

$$r_k = \frac{\Delta f_k}{\Delta p_k}. \tag{2.2}$$

文中利用文献 [10]中所给的信赖域算法与 (1.3) 相结合给出一个新的 BFGS校正的信赖域算法.

新的 BFGS信赖域算法

步骤 0 给定 $x_0 \in R^n$,对称阵 $B_0 \in R^{n \times n}$, $\Delta_0 > 0$;

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 (2.1),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 $\Delta_{k+1} = \|s_k\|/4$;如果 $r_k > 0.75$和 $\|s_k\| = \Delta_k$,令 $\Delta_{k+1} = 2\Delta_k$;否则置 $\Delta_{k+1} = \Delta_k$;

步骤 5 若 $r_k \leqslant 0$,置 $x_{k+1} = x_k + s_k$;否则置 $x_{k+1} = x_k + s_k$,利用校正公式 (1.3)产生 B_{k+1},令 $k := k+1$,转步骤 1.

2003-12-04收稿, 2004-04-15修回。
* 广西自然科学基金资助项目 (9811020).

由以上算法可得到下面的性质.

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

$$y_k^{*T} s_k = \frac{y_k^T s_k}{|y_k^T s_k|} y_k^T s_k = \frac{(y_k^T s_k)^2}{|y_k^T s_k|} = |s_k^T B_k s_k + o(\| s_k \|^2)| \geqslant |V_k + o(1)| \| s_k \|^2 > 0, \quad (2.3)$$

其中 $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$; $s_k = x_{k+1} - x_k$, V_k是 B_k的最小特征值.所以 B_{k+1}是正定矩阵.当 B_0取为对称正定矩阵时,可用数学归纳法证得由校正公式(1.3)产生的 B_k是正定的.

3 全局收敛性和收敛速度

为建立新算法的全局收敛性,现先作一些假设条件,首先,定义水平集

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

假设条件为:

(A): f在 $L(x_0)$是二次连续可微的,且存在 $M > 0$,使得 $\| B_k \| < M$成立.

(B): 存在 $m > 0$使得

$$t^T \nabla^2 f(x) t \geqslant m \| t \|^2, \forall x \in L(x_0), t \in R^n.$$

注 2 根据注 1的讨论知 B_k是正定的,所以假设(B)是合理的.根据文献[10]中的定理 1.3.16可得到水平集 $L(x_0)$是有界闭凸集,所以$\{x_k\}$存在聚点.设 $x_k \to x^*$,令 $B^* = \nabla^2 f(x^*)$,则可推得 B^*是正定的.

由假设(A)和(B)可得到新算法是全局收敛的,下面给出算法 1的全局收敛性定理.

定理 1 由假设(A),(B)以及上述讨论可知,新算法 1产生一个满足一阶和二阶必要条件的聚点 x^*.

证明 由算法产生的序列中存在一个子序列,满足下列情况之一

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

(2) $r_k \geqslant 0.25$, $glb(Z_k) > 0$(glb表示 Z_k的总体最小界).

取 x^*为上述子序列的任一聚点.

首先讨论情形(1).反证法,假设 $g(x^*) \neq 0$,对任何 x_k,考虑最速下降步,可有

$$p_k\left(-\frac{U g_k}{\| g_k \|_2}\right) = f_k - U \| g_k \|_2 + \frac{1}{2} U^2 \frac{g_k^T B_k g_k}{\| g_k \|_2^2}, \quad (3.1)$$

其中,U是搜索步长.

若 $g_k^T B_k g_k > 0$,则当

$$U_{\min} = \frac{\| g_k \|_2^3}{g_k^T B_k g_k}$$

时,$p_k\left(-\frac{U g_k}{\| g_k \|_2}\right)$取得极小值,记为

$$d_{\min} = \frac{1}{2} \frac{\| g_k \|_2^4}{g_k^T B_k g_k},$$

由于 $s = -\frac{Z_k g_k}{\| g_k \|_2}$对于信赖域子问题(2.1)是可行的,所以

$$\Delta p_k = f_k - p_k(s_k) \geqslant f_k - p_k(s) = f_k - p_k\left(-\frac{Z_k g_k}{\| g_k \|_2^2}\right) = Z_k \| g_k \|_2 - \frac{1}{2} \frac{Z_k^2 g_k^T B_k g_k}{\| g_k \|_2^2} = \frac{1}{2} Z_k \| g_k \|_2 \left(2 - Z_k \frac{g_k^T B_k g_k}{\| g_k \|_2^3}\right) = \frac{d_{\min} Z_k}{U_{\min}}\left(2 - \frac{Z_k}{U_{\min}}\right), \quad (3.2)$$

因为 $Z_k \to 0$, $U_{min} \geqslant \frac{\| g_k \|_2}{M} \to \frac{\| g(x^*) \|_2}{M} > 0$,所以对充分大的 k,可得

$$\Delta p_k \geqslant \frac{d_{\min} Z_k}{U_{\min}} = \frac{1}{2} Z_k \| g_k \|_2. \quad (3.3)$$

如果 $g_k^T B_k g_k < 0$,从(3.2)中第三个等式也可直接得出这一结果.因此,当 $\| s_k \| \to 0$时,

$$\frac{\| s_k \|_2^2}{\Delta p_k} \leqslant \frac{\| s_k \|_2^2}{Z_k \| g_k \|_2} \leqslant \frac{2}{\| g_k \|_2} \to 0, \quad (3.4)$$

再由泰勒展开式,

$$\Delta f_k = \Delta p_k + o(\| s_k \|_2^2),$$

所以得到

$$r_k = \frac{\Delta f_k}{\Delta p_k} \to 1.$$

这与 $r_k < 0.25$矛盾,所以 $g(x^*) = 0$成立.

若 B^*是半正定的,现在讨论第二种情形(2).注意到

$$f_k - f(x^*) \geqslant \sum_k \Delta f_k,$$

再根据 $f_1 - f(x^*)$是常数,所以由

$$\sum_k \Delta f_k \geqslant 0.25 \sum_k \Delta p_k,$$

因此可得到,$\Delta p_k \to 0$.文中定义

$$p^*(s) = f(x^*) + s^T g(x^*) + \frac{1}{2} s^T B^* s,$$

设 Z'满足 $0 < Z' < glb(Z_k)$,又设 s'在约束条件 $\| s \|_2 \leqslant Z'$之下极小化 $p^*(s)$.对充分大的 k, $x' = x^* + s'$关于 $v_k = \{x \| x - x_k \| \leqslant Z_k\}$可行,这样

$$p_k(x' - x_k) \geqslant p_k(s_k) = f_k - \Delta p_k,$$

取极值有,$x' - x_k \to s'$, $f_k \to f(x^*)$, $B_k \to B^*$, $g_k \to g(x^*)$, $\Delta p_k \to 0$,从而得到

$$p^*(s') \geqslant f(x^*) = p^*(0),$$

因此 $s = 0$在约束条件 $\| s \| \leqslant Z'$之下也极小化 $p^*(s)$.注意到 $Z' > 0$,故这时约束 $\| s \| \leqslant Z'$是无效约束,其对应的拉格朗日成子为零,从而约束问题

(下转第 200页 Continue on page 200)

$$\sum_{k=0}^{\infty}\left(\frac{s_k^T B_k s_k\|g_k\|}{\|B_k s_k\|}\right)=\sum_{k=0}^{\infty}\left(T_k\|g_k\|^2\frac{s_k^T B_k s_k}{\|B_k s_k\|^2}\right)\geqslant$$

$$X\sum_{k=0}^{\infty}\left(T_k\frac{s_k^T B_k s_k}{\|B_k s_k\|^2}\right).$$

所以对任意的 $l>0$ 存在常数 k_0 满足任意的正整数 p,

$$p\left[\prod_{k=k_0+1}^{k_0+p}T_k\frac{s_k^T B_k s_k}{\|B_k s_k\|^2}\right]^{\frac{1}{p}}\leqslant\sum_{k=k_0+1}^{k_0+p}\frac{s_k^T B_k s_k}{\|B_k s_k\|^2}\leqslant l,$$

上式左边的不等式利用了算术平均与几何平均不等式的关系.因此

$$\left(\prod_{k=k_0+1}^{k_0+p}T_k\right)^{\frac{1}{p}}\leqslant\frac{l}{p}\left(\prod_{k=k_0+1}^{k_0+p}\frac{\|B_k s_k\|^2}{s_k^T B_k s_k}\right)^{\frac{1}{p}}\leqslant$$

$$\frac{l}{p}\sum_{k=k_0+1}^{k_0+p}\frac{\|B_k s_k\|^2}{s_k^T B_k s_k}\leqslant\frac{l}{p}\sum_{k=0}^{k_0+p}\frac{\|B_k s_k\|^2}{s_k^T B_k s_k}$$

$$\leqslant\frac{l(k_0+p+1)}{p^2}M_1,$$

令 $p\to\infty$ 将产生一个矛盾,因为根据引理 3.3上式不等式的左边是大于一个正常数的.

所以假设不成立,原命题成立.

3 结束语

本文给出了一个修改的 BFGS方法,此方法让 y_k 乘上了 $y_k^T s_k$ 的符号,得到了与通常的 BFGS方法同样的性质,特别是保持正定性.可将此种方法推广至其它方面做进一步的研究,应会得到较好的结果.对于此方法的超线性收敛性见文献 [10]的第 5.5节类似的证明,这里不再论证.

参考文献

1 Dennis J E, Schnabel R B. Numerical methods for unconstrained optimization and nonlinear equations. Englewood Cliffs: Pretice-Hall Inc, 1983.

2 Y Dai. Convergence properties of the BFGS algorithm. SIAM Journal on Optimization, 2003, (13): 693～701.

3 Fletcher R. Practical methods of optimization. 2nd ed. Chichester: John Wiley & Sons, 1987.

4 Griewank A, Ph L Toint. Local convergence analysis for partitioned quasi-Newton updates. Numer Math, 1982, (39): 429～448.

5 Davidon W C. Variable metric methods for minimization. Argonne National Labs Report, ANL-5990. 2000.

6 Li D, Fukushima M. A modified BFGS method and its global convergence in nonconvex minimization. Journal of Computational and Applied Mathematics, 2001, (129): 15～35.

7 Powell M J D. A new algorithm for unconstrained optimization. In: Nonlinear Programming. Rosen J B, Mangasarian O L, Ritter K. eds. New York: Academic Press, 1970.

8 Wei Z, Qi L, Chen X. A SQP-typemethod and its application in stochastic programming. Journal of Optimization Theory and Applications, 2003, (116): 205～228.

9 Dennis J E JR, Moré J J. Quasi-Newton methods motivation and theory. SIAM Rev, 1977, (19): 46～89.

10 Yuan Y, Sun W. Theory and Methods of Optimization. Beijing: Science Press of China. 1999.

(责任编辑:黎贞崇)

(上接第 196页 Continue from page 196)

$$\min p^*(s),$$
$$\text{s.t.}\ \|s\|\leqslant Z$$

的一阶必要条件简化为 $g(x^*)=0$,二阶必要条件简化为 B^* 半正定.证毕.

参考文献

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

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

3 Byrd R H, Schnabel R B, Shultz G A. Approximate solution of the trust region problem by minimization over two-dimensional subspaces. Math Prog, 1988, 40: 247～263.

4 袁亚湘.信赖域方法的收敛性.计算数学, 1994, 16: 333～346.

5 Dennis J E, Schnabel R B. Numerical methods for unconstrained optimization and nonlinear equations. Englewood Cliffs: Pretice-Hall Inc, 1983.

6 Fletcher R. Practical meethods of optimization. 2nd ed. Chichester: John Wiley & Sons, 1987.

7 Dennis J E, Jr Moré J J. A characteization of superlinear convergence and its application to Quasi-Newton methods. Math Comp, 1974, 28: 1171～1190.

8 Broyden C G, Dennis J E, Moré J J. On the local and supelinear convergence of Quasi-Newton methods. J Inst Math Appl, 1973, 12: 223～246.

9 Byrd R, Nocedal J, Yuan Y. Global convergence of a class of Quasi-Newton methods on convex problems. SIAM Journal on Numerical Analysis, 1987, 24: 1171～1189.

10 Yuan Y, Sun W. Theory and Methods of Optimization. Beijing: Science Press of China, 1999.

(责任编辑:黎贞崇)