DOI: 10.13656/j.cnki.gxkx.2008.03.015
广西科学 Guangxi Sciences 2008, 15(3): 254- 256

曲线搜索下新的记忆拟牛顿算法*

A New Memory Quasi-Newton Method with Curve Search Rule

陈凤华,张 聪,房明磊

CHEN Feng-hua, ZHANG Cong, FANG Ming-lei

(桂林电子科技大学数学与计算科学学院,广西桂林 541004)

(School of Mathematics and Computational Science, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China)

摘要: 利用新的曲线搜索方法,提出一种解决无约束优化问题的记忆拟牛顿算法,给出该算法全局收敛的条件并进行数值实验.新算法由曲线搜索确定迭代步长,搜索方向用到当前迭代点信息的同时还用到上一次迭代点的信息,而且搜索方向与迭代步长同时确定,是一种有效的算法.

关键词: 无约束优化 记忆拟牛顿算法 全局收敛 曲线搜索
中图法分类号: O221.2 文献标识码: A 文章编号: 1005-9164(2008)03-0254-03

Abstract By means of a new curve search rule, we propose a new memory quasi-Newton method for unconstrained optimization problems. A global convergent result is proved and numerical experiments are carried out. The step-size of the new algorithm is determined by the new curve search rule. The search direction not only make use of the current iterative information but also the previous iterative information. Moreover, the search direction and the step-size are determined simultaneously at each iteration. Numerical results show that the algorithm is effective.

Key words unconstrained optimization, memory quasi-Newton method, global convergence, curve search rule

考虑优化问题

$$\min f(x), x \in R^n, \quad (1)$$

其中 $f: R^n \to R$ 为连续可微函数.求解问题(1)的迭代形式为

$$x^{k+1} = x^k + T_k d^k, k = 1, 2, \cdots, \quad (2)$$

其中 d^k 是搜索方向,T_k 是迭代步长.为了方便,记 $\nabla f(x^k)$ 为 g^k, $f(x^k)$ 为 f_k, $f(x^*)$ 为 f^*.

在(2)式中,若 $d^k = -B_k^{-1} g^k$, B_k 是 Hessian 阵 $\nabla^2 f(x^k)$ 的近似,则相关的下降算法是拟牛顿算法[1].通常的算法要求搜索方向 d^k 满足

$$g^{kT} d^k < 0, \quad (3)$$

这样才能保证 d^k 是下降方向.

记 $r^k = x^{k+1} - x^k$, $y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$,若 $r^{kT} y^k$ 充分小,那么用 BFGS形式、DFP形式或其它拟牛顿形式修正 B_k 为 B_{k+1}, B_{k+1} 满足拟牛顿条件 $B_{k+1} y_k = r^k$.

传统的线搜索下的拟牛顿算法[1]是先确定搜索方向,搜索方向由 $d^k = -B_k^{-1} g^k$ 确定,然后沿着搜索方向由线搜索确定步长,搜索方向 d^k 与迭代步长 T_k 无关.本文提出的曲线搜索下的新拟牛顿算法的特点是:(I)搜索方向是迭代步长的向量函数,搜索方向和迭代步长是同时确定的;(II)搜索方向用到当前迭代点信息的同时,还用到上一次迭代点的信息;(III)由曲线搜索确定迭代步长.

1 算法描述

对问题(1)作基本假设:

H1 函数 f 在 L_0 上有下界,其中 $L_0 = \{x \in R^n | f(x) \leqslant f(x^1)\}$, x^1 给定.

H2 梯度函数 g 在开凸集 B 上 Lipschitz 连续,即存在常数 $L > 0$ 满足

$$\|g(x) - g(y)\| \leqslant L\|x - y\|, \forall x, y \in B. \quad (4)$$

新算法的曲线搜索方法:在每步迭代中,步长 T_k 满足

收稿日期: 2007-07-18
修回日期: 2008-04-10
作者简介: 陈凤华(1982-),女,硕士研究生,主要从事最优化理论与算法研究工作。
* 国家自然科学基金项目(10501009),中国博士后基金项目(20070410227)与广西自然科学基金项目(桂科自 0728206)资助。

$$f(x^k+\alpha d^k(\alpha))-f_k\leqslant\sigma_1\alpha g^{kT}d^k(\alpha) \qquad (5)$$

和

$$g^T(x^k+\alpha d^k(\alpha))d^k(\alpha)\geqslant\sigma_2 g^{kT}d^k(\alpha), \qquad (6)$$

其中 $0<\sigma_1<\frac{1}{2}<\sigma_2<1$.

定义

$$s_k=\begin{cases}1, & k=1,\\ \dfrac{g^{kT}B_k^{-1}g^k}{g^{kT}B_k^{-1}g^k+|g^{kT}d^{k-1}|}, & k\geqslant 2.\end{cases} \qquad (7)$$

构造搜索方向:

$$d^k(\alpha_k)=\begin{cases}-B_k^{-1}g^k, & k=1,\\ -\left[\left(1-\dfrac{\alpha_k s_k}{1+\alpha_k}\right)B_k^{-1}g^k+\dfrac{\alpha_k s_k}{1+\alpha_k}d^{k-1}\right], & k\geqslant 2.\end{cases} \qquad (8)$$

为了方便,记 $d^k(\alpha_k)$ 为 d^k. 由(8)式可以知道, $k\geqslant 2$ 时,搜索方向 d^k 是迭代步长 α_k 的向量函数. 把(7)式和(8)式代入(5)式和(6)式得到关于 α_k 的不等式,求得 α_k, α_k 代入(8)式,确定 d^k,即搜索方向 d^k 和迭代步长 α_k 是同时确定的.

算法 1

步骤 1: 给定 x^1 和 B_1, $x^1\in R^n$, B_1 为单位矩阵,置 $k:=1$.

步骤 2 若 $\|g^k\|=0$ 停止! 否则转步骤 3.

步骤 3 按(7)式和(8)式确定搜索方向,按(5)式和(6)式确定搜索步长,置 $x^{k+1}=x^k+\alpha_k d^k(\alpha_k)$.

步骤 4 令 $k:=k+1$, 转步骤 2.

2 算法的收敛性分析

引理 1 对任何 $k\geqslant 2$, 有 $-g^{kT}d^k\geqslant\dfrac{g^{kT}B_k^{-1}g^k}{1+\alpha_k}$.

证明 当 $k\geqslant 2$ 时,

$$-g^{kT}d^k=-g^{kT}\left[-\left(1-\frac{\alpha_k s_k}{1+\alpha_k}\right)B_k^{-1}g^k-\frac{\alpha_k s_k}{1+\alpha_k}d^{k-1}\right]=\left(1-\frac{\alpha_k s_k}{1+\alpha_k}\right)g^{kT}B_k^{-1}g^k+\frac{\alpha_k s_k}{1+\alpha_k}g^{kT}d^{k-1}=$$

$$g^{kT}B_k^{-1}g^k-\frac{\alpha_k s_k}{1+\alpha_k}(g^{kT}B_k^{-1}g^k-g^{kT}d^{k-1})\geqslant g^{kT}B_k^{-1}g^k-\frac{\alpha_k s_k}{1+\alpha_k}(g^{kT}B_k^{-1}g^k+|g^{kT}d^{k-1}|)=g^{kT}B_k^{-1}g^k-\frac{\alpha_k}{1+\alpha_k}\cdot g^{kT}B_k^{-1}g^k=\frac{g^{kT}B_k^{-1}g^k}{1+\alpha_k}.$$

引理 2 对任何 $k\geqslant 2$, 有 $\|d^k\|^2\leqslant\max\{\|B_k^{-1}g^k\|^2,\|d^{k-1}\|^2\}\leqslant\|B_k^{-1}g^k\|^2+\|d^{k-1}\|^2$.

证明 当 $k\geqslant 2$ 时,

$$\|d^k\|^2=\left[\left(1-\frac{\alpha_k s_k}{1+\alpha_k}\right)B_k^{-1}g^k+\frac{\alpha_k s_k}{1+\alpha_k}d^{k-1}\right]^T\left[\left(1-\frac{\alpha_k s_k}{1+\alpha_k}\right)B_k^{-1}g^k+\frac{\alpha_k s_k}{1+\alpha_k}d^{k-1}\right]=\left(1-\frac{\alpha_k s_k}{1+\alpha_k}\right)^2\|B_k^{-1}g^k\|^2+2\left(1-\frac{\alpha_k s_k}{1+\alpha_k}\right)\frac{\alpha_k s_k}{1+\alpha_k}g^{kT}B_k^{-1}d^{k-1}+\left(\frac{\alpha_k s_k}{1+\alpha_k}\right)^2\|d^{k-1}\|^2\leqslant$$

$$\left(1-\frac{\alpha_k s_k}{1+\alpha_k}\right)^2\max\{\|B_k^{-1}g^k\|^2,\|d^{k-1}\|^2\}+2\left(1-\frac{\alpha_k s_k}{1+\alpha_k}\right)\frac{\alpha_k s_k}{1+\alpha_k}\max\{\|B_k^{-1}g^k\|^2,\|d^{k-1}\|^2\}+\left(\frac{\alpha_k s_k}{1+\alpha_k}\right)^2\max\{\|B_k^{-1}g^k\|^2,\|d^{k-1}\|^2\}=\max\{\|B_k^{-1}g^k\|^2,\|d^{k-1}\|^2\},$$

即 $\|d^k\|^2\leqslant\max\{\|B_k^{-1}g^k\|^2,\|d^{k-1}\|^2\}\leqslant\|B_k^{-1}g^k\|^2+\|d^{k-1}\|^2$.

推论 1 对任何 $k\geqslant 2$, 有 $\|d^k\|^2\leqslant\max\limits_{1\leqslant i\leqslant k}\{\|B_i^{-1}g^i\|^2\}\leqslant\sum_{i=1}^{k}\|B_i^{-1}g^i\|^2$.

证明 由数学归纳法知,当 $k=2$ 时,

$\|d^2\|^2\leqslant\max\{\|B_2^{-1}g^2\|^2,\|d^1\|^2\}=\max\{\|B_2^{-1}g^2\|^2,\|B_1^{-1}g^1\|^2\}=\max\limits_{1\leqslant i\leqslant 2}\|B_i^{-1}g^i\|^2$.

假设 $k=n$ 时结论成立,即 $\|d^n\|^2\leqslant\max\limits_{1\leqslant i\leqslant n}\{\|B_i^{-1}g^i\|^2\}$.

事实上,

$\|d^{n+1}\|^2\leqslant\max\{\|B_{n+1}^{-1}g^{n+1}\|^2,\|d^n\|^2\}\leqslant\max\{\|B_{n+1}^{-1}g^{n+1}\|^2,\max\limits_{1\leqslant i\leqslant n}\{\|B_i^{-1}g^i\|^2\}\}=\max\limits_{1\leqslant i\leqslant n+1}\|B_i^{-1}g^i\|^2$,即 $k=n+1$ 时,结论亦成立. 所以结论得证.

引理 3 设 H1, H2 成立,则 $\alpha_k\geqslant(-(1-\sigma_2)g^{kT}d^k(\alpha_k))/L\|d^k(\alpha_k)\|^2$.

证明 由 H1, Cauchy-Schwarz 不等式和(6)式可以得到

$$\alpha_k L\|d^k(\alpha_k)\|^2\geqslant\|g(x^k+\alpha_k\theta_k d^k(\alpha_k))-g_k\|\,\|d^k(\alpha_k)\|\geqslant[g(x^k+\alpha_k\theta_k d^k(\alpha_k))-g_k]^T d^k(\alpha_k)\geqslant-(1-\sigma_2)g^{kT}d^k(\alpha_k),$$

即 $\alpha_k\geqslant-\dfrac{(1-\sigma_2)g^{kT}d^k(\alpha_k)}{L\|d^k(\alpha_k)\|^2}$.

定理 1 设 H1, H2 成立,且算法产生一无穷序列 $\{x^k\}$,则

$$\sum_{k=2}^{\infty}\frac{(g^{kT}B_k^{-1}g^k)^2}{(1+\alpha_k)^2\max\limits_{1\leqslant i\leqslant k}\|B_i^{-1}g^i\|^2}<+\infty. \qquad (9)$$

证明 由(5)式,引理 3,推论 1 和引理 1,有

$$f_k-f_{k+1}\geqslant-\alpha_k\sigma_1 g^{kT}d^k(\alpha_k)\geqslant\frac{\sigma_1(1-\sigma_2)}{L}\cdot\frac{(g^{kT}d^k(\alpha_k))^2}{\|d^k(\alpha_k)\|^2}\geqslant\frac{\sigma_1(1-\sigma_2)}{L}\cdot\frac{(g^{kT}d^k(\alpha_k))^2}{\max\limits_{1\leqslant i\leqslant k}\|B_i^{-1}g^i\|^2}\geqslant\frac{\sigma_1(1-\sigma_2)}{L}\cdot\frac{(g^{kT}B_k^{-1}g^k)^2}{(1+\alpha_k)^2\max\limits_{1\leqslant i\leqslant k}\|B_i^{-1}g^i\|^2}.$$

因为 $\{f_k\}$ 是一下降序列且有下界,则 $\{f_k\}$ 极限存在,所以(9)式成立.

再给出 2 个假设条件:

H3 $\sup\limits_{\forall k}\{\alpha_k\}<+\infty$.

H4 矩阵 B_k 满足 $m\|d\|^2\leqslant d^T B_k d\leqslant M\|d\|^2$, $\forall d\in R^n$, $\forall k$,其中 $0<m\leqslant M$.

推论 2 设 H1, H2, H3 成立,且算法产生一无穷序列 $\{x^k\}$,则

$$\sum_{k=2}^{\infty}\frac{(g^{kT}B_k^{-1}g^k)^2}{\max\limits_{1\leqslant i\leqslant k}\|B_i^{-1}g^i\|^2}<+\infty. \qquad (10)$$

引理 4 设 H1, H2, H3, H4 成立, 且算法产生一无穷序列 $\{x^k\}$, 则序列 $\{\|B_k^{-1}g^k\|\}$ 和 $\{\|d^k(\tau_k)\|\}$ 均有界, 记上界为 M_0.

证明 令 $\varepsilon_k = \max\limits_{1\le i\le k}\{\|B_i^{-1}g^i\|^2\}$. 假设 $\{\|B_k^{-1}g^k\|\}$ 无界, 则存在一无穷集 N, 满足

$$\|B_k^{-1}g^k\|^2 = \varepsilon_k \to +\infty, k \to +\infty. \qquad (11)$$

又由 H4, 有 $m\|B_k^{-1}g^k\|^2 \leqslant g^{kT}B_k^{-1}g^k = (B_k^{-1}g^k)^T B_k (B_k^{-1}g^k)$, 因此,

$$\sum_{k\in N}\|B_k^{-1}g^k\|^2 = \sum_{k\in N}\frac{\|B_k^{-1}g^k\|^4}{\|B_k^{-1}g^k\|^2} = \sum_{k\in N}\frac{\|B_k^{-1}g^k\|^4}{\varepsilon_k} \leqslant \sum_{k=2}^{\infty}\frac{\|B_k^{-1}g^k\|^4}{\varepsilon_k} \leqslant$$

$$\frac{1}{m^2}\sum_{k=2}^{\infty}\frac{(g^{kT}B_k^{-1}g^k)^2}{\max\limits_{1\le i\le k}\|B_i^{-1}g^i\|^2} < +\infty,$$

这与 (11) 式矛盾, 则 $\{\|B_k^{-1}g^k\|\}$ 有界. 进一步地由推论 1 可知, $\{\|d^k(\tau_k)\|\}$ 亦有界.

定理 2 设 H1~ H4 成立, 且算法产生一无穷序列 $\{x^k\}$, 则 $\lim\limits_{k\to\infty} g^{kT}B_k^{-1}g^k = 0$.

证明 由引理 4, 有

$$\|d^k(\tau_k)\|^2 \leqslant \max_{1\le i\le k}\{\|B_i^{-1}g^i\|^2\} \leqslant M_0^2, \quad \forall k.$$

又由推论 2, 有

$$\frac{1}{M_0^2}\sum_{k=2}^{\infty}(g^{kT}B_k^{-1}g^k)^2 \leqslant \sum_{k=2}^{\infty}\frac{(g^{kT}B_k^{-1}g^k)^2}{\max\limits_{1\le i\le k}\|B_i^{-1}g^i\|^2} < +\infty.$$

因此, $(g^{kT}B_k^{-1}g^k)^2 \to 0 (k \to +\infty)$, 即 $g^{kT}B_k^{-1}g^k \to 0 (k \to +\infty)$.

引理 5 若 H4 成立, 则 $\frac{1}{M}\|d\|^2 \leqslant d^T B_k^{-1} d \leqslant \frac{1}{m}\|d\|^2, \forall k$.

证明 因 B_k 是一对称正定阵, 则存在一正交阵 $p \in R^n$, 满足

$$m\|d\|^2 \leqslant d^T p^T \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} p d \leqslant M\|d\|^2,$$

其中 $\lambda_i (i = 1, 2, \cdots, n)$ 为 B_k 的特征根.

在不等式两边取逆, 有

$$\frac{1}{M}\|d\|^2 \leqslant d^T p^T \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}^{-1} p d \leqslant \frac{1}{m}\|d\|^2,$$

即 $\frac{1}{M}\|d\|^2 \leqslant d^T B_k^{-1} d \leqslant \frac{1}{m}\|d\|^2, \forall k$.

定理 3 设 H1~ H4 成立, 且算法产生一无穷序列 $\{x^k\}$, 则 $\lim\limits_{k\to\infty}\|g^k\| = 0$.

证明 由引理 5, 有 $\frac{1}{M}\|g^k\|^2 \leqslant g^{kT}B_k^{-1}g^k, \forall k$.

又由定理 2, 有 $\lim\limits_{k\to\infty}\|g^k\|^2 = 0$, 即 $\lim\limits_{k\to\infty}\|g^k\| = 0$.

3 数值实验

通过 Matlab 编程将算法在计算机上进行数值实验, 并与文献 [1~ 4] 的算法进行比较, 所得结果见表 1 和表 2.

问题 **1**[3,4] $\min f(x) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$. 设 $_1 = 0.38$, $_2 = 0.85$, 精度 $p < 10^{-10}$. $B_1 = I$, B_k 由 BFGS 公式修正: $B_{k+1} = \mathrm{BFGS}(B_k, r^k, y^k)$, 其中 $r^k = x^{k+1} - x^k$, $y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$.

表 1 关于问题 1 的算法比较

Table 1 Compare algorithm on problem 1

算法 Algorithm	初始点 Initial point	迭代次数 Iterative time(k)	x	$\|g^k\|$
算法 1 Algorithm 1	(1, 1)	32	(3.0000000000100, 2.0000000000199)	6.76296293952309e-011
算法 2[3] Algorithm 2	(1, 1)	20	(2.99999999999875, 2.0000000000199)	7.959871e-011
算法 3[4] Algorithm 3	(1, 1)	60	(3.0000, 2.0000)	7.1684e-011

问题 **2**[2] $\min f(x) = (x_1 + 10 * x_2)^4 + 5 * (x_3 - x_4)^4 + (x_2 - 2 * x_3)^4 + 10 * (x_1 - x_4)^4$. 设 $_1 = 0.2$, $_2 = 0.85$, 精度 $p < 10^{-8}$.

表 2 关于问题 2 的算法比较

Table 2 Compare algorithm on problem 2

算法 Algorithm	初始点 Initial point	k/k^*	x	$\|g^k\|$
算法 1 Algorithm	(2, 2, - 2, - 2)	117/214	(- 0.00039701077925842 2, 8.9747783586077 3e- 005, 0.000125822239892798, - 6.34594952893687e- 005)	5.2711186602246 9e- 009
NM 算法 [2] NM- algorithm	(2, 2, - 2, - 2)	16/227	(0, 0, 0, 0)	
FR-R 算法 [2] NM- algorithm	(2, 2, - 2, - 2)	42/1856	(0, 0, 0, 0)	

* : 迭代次数/目标函数的估计次数. Iterative time/Estimate time of objective function.

表 1 和表 2 的结果表明算法 1 是有效的.

参考文献:

[1] Z J shi. Convergence of quasi-Newton method with new inexact linesearch [J]. J Math Anal, 2006, 315: 120-131.

[2] Z J shi. A new descent algorithm with curve search rule [J]. Comput Math Appl, 2005, 161: 753-768.

[3] 柳娟, 谢铁军, 孙玉华. 一类共轭梯度法的全局收敛性 [J]. 运筹与管理, 2006, 15(3): 75-79.

[4] 房明磊, 张聪, 陈凤华. 一种新的 Wolfe 线搜索技术及全局收敛性 [J]. 桂林电子科技大学学报, 2008, 28(1): 63-65.

(责任编辑: 尹 闯)