DOI:10.13656/j.cnki.gxkx.2005.04.012
广西科学 Guangxi Sciences 2005, 12(4): 282~283, 287

一种新的非线性共轭梯度法的全局收敛性

Global Convergence Properties of a New Class of Nonlinear Conjugate Gradient Methods

张秀军,徐安农
Zhang Xiujun, Xu Annong

(桂林电子工业学院计算科学与数学系,广西桂林 541004)
(Department of Computer Science and Mathematics, Guilin University of Electron Technology, Guilin, Guangxi, 541004, China)

摘要:基于标准 Wolfe线搜索条件,提出一种新的线搜索:α_k满足

$f(x_k+\alpha_k d_k)-f(x_k)\leqslant \max\{\delta\alpha_k g_k^T d_k, -\mu\alpha_k^2\|d_k\|^2\}$ 和 $g(x_k+\alpha_k d_k)^T d_k\geqslant \max\{\sigma g_k^T d_k, -2\sigma\alpha_k\|d_k\|^2\}$,

并在此基础上给出了一种新的非线性共轭梯度算法及其全局收敛性定理.

关键词:无约束优化 共轭梯度 全局收敛性 Wolfe线搜索
中图法分类号:O224 文献标识码:A 文章编号:1005-9164(2005)04-0282-02

Abstract A new line search based on the standard Wolfe line search conditions α_k satisfing

$f(x_k+\alpha_k d_k)-f(x_k)\leqslant \max\{\delta\alpha_k g_k^T d_k, -\mu\alpha_k^2\|d_k\|^2\}$ and $g(x_k+\alpha_k d_k)^T d_k\geqslant \max\{\sigma g_k^T d_k, -2\sigma\alpha_k\|d_k\|^2\}$, was present in this paper, and the global convergence of a new class of nonlinear conjugate gradient methods with the new line search was given.

Key words unconstrained optimization, conjugate gradient, global convergence, Wolf line search

考虑无约束优化问题

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

其中,$f: R^n\to R^1$连续可微,其梯度向量记为 g.求解的有效方法是非线性共轭梯度方法,其迭代公式为:

$$x_{k+1}=x_k+\alpha_k d_k, \tag{2}$$

$$d_k=\begin{cases}-g_k, & k=1,\\ -g_k+\beta_{k-1}d_{k-1}, & k\geqslant 2,\end{cases} \tag{3}$$

其中,α_k为步长因子;β_k为参数.关于β_k计算公式有很多,其中较著名的有 Fletcher-Reeves(FR)[1], Polak-Ribiere(PR)[2], Hestenes-Stiefel(HS)[3], Conjugate-Decent(CD)[4]公式.文献[5]提出了一种新的β_k计算公式,即

$$\beta_k^{DY}=\|g_{k+1}\|^2/d_k^T y_k, \tag{4}$$

并证明了这类共轭梯度法在标准 Wolf线搜索下的全局收敛性.文献[6]减弱了搜索条件而得到一种新的 DY共轭梯度法,并证明了算法的全局收敛性.文献[7]给出了一种新的线搜索,并证明了一般共轭梯度法的全局收敛性.本文结合文献[6~8]提出了一种新的线搜索,并在此基础上给出了一种新的非线性共轭梯度算法及其全局收敛性定理.

1 非线性共轭梯度算法

标准 Wolf线搜索条件为:

$$f(x_k+\alpha_k d_k)-f(x_k)\leqslant \delta\alpha_k g_k^T d_k, \tag{5}$$

$$g(x_k+\alpha_k d_k)^T d_k\geqslant \sigma g_k^T d_k, \tag{6}$$

其中,$0<\delta<\sigma<1$.文献[5]证明了 DY共轭梯度法在标准 Wolfe条件下的全局收敛性.

本文基于标准 Wolfe线搜索条件,结合文献[6~8]提出一种新的线搜索:α_k满足

$$f(x_k+\alpha_k d_k)-f(x_k)\leqslant \max\{\delta\alpha_k g_k^T d_k, -\mu\alpha_k^2\|d_k\|^2\}, \tag{7}$$

$$g(x_k+\alpha_k d_k)^T d_k\geqslant \max\{\sigma g_k^T d_k, -2\sigma\alpha_k\|d_k\|^2\}, \tag{8}$$

其中,$0<\delta<\sigma<1$, $0<\mu<1$.

新的线搜索条件(7),(8)与标准 Wolfe线搜索条件(5),(6)相比可知:(7)较(5)能保证$f(x)$下降的更大,(8)较(6)更能防止步长过小,因而保证目标函数的足够下降.所以新的线搜索(α_k满足(7),(8))比

收稿日期:2005-06-06
修回日期:2005-09-06
作者简介:张秀军(1979-),男,山东临沂人,硕士研究生,主要从事最优化理论与算法研究

标准 Wolfe线搜索(α_k满足(5),(6))以及文献[6,7]中的线搜索都要好.

根据新的线搜索条件,我们提出如下的新算法:

步骤 1:取 $x_1 \in R^n$, $d_1 = -g_1$, $k = 1$.若 $g_1 = 0$,则停;

步骤 2 计算 $\alpha_k > 0$,满足条件(7),(8);

步骤 3 令 $x_{k+1} = x_k + \alpha_k d_k$.若 $g_{k+1} = 0$,则停;

步骤 4 计算 $\beta_k^{PY} = \|g_{k+1}\|^2 / d_k^T y_k$,由(3)式得 d_{k+1}, $k = k+1$.转步骤 2.

由(3)和(4)式可以得到

$$g_{k+1}^T d_{k+1} = g_{k+1}^T(-g_{k+1} + \beta_k^{PY} d_k) =$$

$$\|g_{k+1}\|^2 \frac{-d_k^T y_k + g_{k+1}^T d_k}{d_k^T y_k} = \frac{\|g_{k+1}\|^2}{d_k^T y_k} g_k^T d_k =$$

$$\beta_k^{PY} g_k^T d_k,$$

从而得到等价形式的公式

$$\beta_k^{PY} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \qquad (9)$$

2 相关引理

为了文后证明的方便,本文总假设目标函数满足以下假设:

假设 **1** (H_1) $f(x)$ 在水平集 $L_0 = \{x \mid f(x) \leqslant f(x_0)\}$ 有界;

(H_2) 在 L 的一个邻域 N 内,f 连续可微,且其梯度向量 g 满足 Lipschitz条件,即存在常数 $L > 0$使得

$$\|g(x) - g(y)\| \leqslant L\|x - y\|, \forall x, y \in N. \qquad (10)$$

引理 **1** 假设条件(H_1),(H_2)成立,考虑方法(2),(3),步长因子 α_k由线搜索(7),(8)确定,则或者 $g_k = 0$对某 k成立,或者

$$g_k^T d_k < 0, \qquad (11)$$

对任意的 $k \geqslant 1$成立.

证明 假设对于所有的 $k \geqslant 1$,均有

$$\|g_k\| > 0, \qquad (12)$$

对(3)式两边与 g_{k+1}作内积,并利用(4)式得

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{d_k^T y_k} g_{k+1}^T d_k =$$

$$\|g_{k+1}\|^2 (-1 + \frac{g_{k+1}^T d_k}{d_k^T y_k}) = \|g_{k+1}\|^2 \frac{g_k^T d_k}{d_k^T y_k}, \qquad (13)$$

因为 $d_1 = -g_1$,显然(11)式对 $k = 1$成立.假设 k时,$g_k^T d_k < 0$成立,由(8)式知

$$d_k^T y_k = d_k^T (g_{k+1} - g_k) \geqslant \sigma g_k^T d_k - g_k^T d_k =$$

$$-(1 - \sigma) g_k^T d_k > 0.$$

从而由(12),(13)式以及假设 $g_k^T d_k < 0$,利用数学归纳法可得 $g_{k+1}^T d_{k+1} < 0$.综上所述,$g_k^T d_k < 0$对任意的 $k \geqslant 1$成立.

引理 **2** 假设条件(H_1),(H_2)成立,考虑方法(2),(3)式,步长因子 α_k由线搜索(7),(8)确定,则

$$\sum_{k \geqslant 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \qquad (14)$$

证明 由引理 1和假设(H_1),可知 $\{f(x_k)\}$为单调下降有界数列,即 $\{f(x_k)\}$为收敛数列.由(8)式可得

$$d_k^T (g_{k+1} - g_k) \geqslant (\sigma - 1) g_k^T d_k, \qquad (15)$$

由假设(H_2)得

$$d_k^T (g_{k+1} - g_k) \leqslant L \alpha_k \|d_k\|^2, \qquad (16)$$

所以由(15),(16)式得

$$\alpha_k \geqslant \frac{\sigma - 1}{L} \frac{g_k^T d_k}{\|d_k\|^2},$$

再由(7)式可知,若 $\rho \alpha_k g_k^T d_k < -\delta \alpha_k^2 \|d_k\|^2$,则有

$$f_k - f_{k+1} \geqslant c_1 \frac{(g_k d_k)^2}{\|d_k\|^2}, c_1 = \delta \frac{(1 - \sigma)^2}{L^2}, \qquad (17)$$

否则,有

$$f_k - f_{k+1} \geqslant c_2 \frac{(g_k d_k)^2}{\|d_k\|^2}, c_2 = \rho \frac{(1 - \sigma)}{L}. \qquad (18)$$

令 $c = \min\{c_1, c_2\}$,则由(17),(18)式得

$$f_k - f_{k+1} \geqslant c \frac{(g_k^T d_k)^2}{\|d_k\|^2}, \qquad (19)$$

由 $\{f(x_k)\}$的收敛性和(19)式得

$$\sum_{k \geqslant 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty.$$

引理得证.

3 全局收敛性

定理 **1** 假设条件(H_1),(H_2)成立,考虑方法(2),(3),步长因子 α_k由线搜索(7),(8)确定,则算法或者终止于稳定点或者

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

证明 若定理不成立,则存在 $\gamma > 0$,使

$$\|g_k\| \geqslant \gamma, \forall k \geqslant 1, \qquad (20)$$

将(3)式写成如下形式

$$d_{k+1} + g_{k+1} = \beta_k d_k, \qquad (21)$$

对(21)式两边平方,得

$$\|d_{k+1}\|^2 = \beta_k^2 \|d_k\|^2 - 2 g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2, \qquad (22)$$

对(22)式两端同除以 $(g_{k+1}^T d_{k+1})^2$,并利用(4)式,得

(下转第 287页 Continue on page 287)

$$E\sum_{j=1}^{n}[I(\frac{|X_j|}{j}>m)-P(\frac{|X_j|}{j}>m)]^2P(\max_{1\leqslant i\leqslant n}\frac{|X_i|}{i}>m)$$

$$\leqslant\frac{1}{2}\sum_{j=1}^{n}P(\frac{|X_j|}{j}>m)\cdot 2cP(\max_{1\leqslant i\leqslant n}\frac{|X_i|}{i}>m)\leqslant\frac{1}{4}\sum_{j=1}^{n}P(\frac{|X_j|}{j}>m)+cp(\max_{1\leqslant i\leqslant n}\frac{|X_i|}{i}>m).$$

由上式及(2.9)、(2.10)式有

$$\frac{3}{4}\sum_{j=1}^{n}P(\frac{|X_j|}{j}>m)\leqslant(1+c)P(\max_{1\leqslant i\leqslant n}\frac{|X_i|}{i}>m)+\sum_{j=1}^{n}P(\frac{|X_j|}{j}>m)P(\max_{1\leqslant i\leqslant n}\frac{|X_i|}{i}>m). \quad (2.11)$$

由 $E\sup_{n\geqslant 1}|\frac{S_n}{n}|<\infty$，有 $E\sup_{n\geqslant 1}|\frac{X_n}{n}|<\infty$. 故当 $m\to\infty$ 时,

$$P(\max_{1\leqslant i\leqslant n}\frac{|X_i|}{i}>m)\leqslant P(\sup_{i\geqslant 1}\frac{|X_i|}{i}>m)\to 0,$$

所以当 m 足够大时有 $P(\max_{1\leqslant i\leqslant n}\frac{|X_i|}{i}>m)\leqslant 1/4$，再结合(2.11)式当 m 足够大时有

$$\sum_{j=1}^{n}P(\frac{|X_j|}{j}>m)\leqslant 2(1+c)P(\max_{1\leqslant i\leqslant n}\frac{|X_i|}{i}>m),$$

令 $n\to\infty$ 有

$$\sum_{j=1}^{\infty}P(\frac{|X_j|}{j}>m)\leqslant 2(1+c)P(\sup_{i\geqslant 1}\frac{|X_i|}{i}>m),$$

从而

$$\sum_{m=1}^{\infty}\sum_{j=1}^{\infty}P(\frac{|X_j|}{j}>m)\leqslant 2(1+c)\sum_{m=1}^{\infty}P(\sup_{i\geqslant 1}\frac{|X_i|}{i}>m)\leqslant 2(c+1)(1+E\sup_{n\geqslant 1}|\frac{X_n}{n}|)<\infty.$$

而由引理知上式正好与 $E|X|\log^+|X|<\infty$ 是等价的. 故而推论得证.

参考文献:

[1] 杨善朝. 混合序列加权和的强收敛性[J]. 系统科学与数学, 1995, 15(3): 254-265.

[2] 严加安. 测度论讲义[M]. 北京: 科学出版社, 1998.

(责任编辑: 黎贞崇)

(上接第283页 Continue from page 283)

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^Td_{k+1})^2}=\frac{\|d_k\|^2}{(g_k^Td_k)^2}-\frac{2}{g_{k+1}^Td_{k+1}}-\frac{\|g_{k+1}\|^2}{(g_{k+1}^Td_{k+1})^2}=\frac{\|d_k\|^2}{(g_k^Td_k)^2}-(\frac{1}{\|g_{k+1}\|}+\frac{\|g_{k+1}\|}{g_{k+1}^Td_{k+1}})^2+\frac{1}{\|g_{k+1}\|^2}\leqslant\frac{\|d_k\|^2}{(g_k^Td_k)^2}+\frac{1}{\|g_{k+1}\|^2}.$$

又因为 $\frac{\|d_1\|^2}{(g_1^Td_1)^2}=\frac{1}{\|g_1\|^2}$，所以

$$\frac{\|d_k\|^2}{(g_k^Td_k)^2}\leqslant\sum_{i=1}^{k}\frac{1}{\|g_i\|^2}, \quad (23)$$

由(20),(23)式可得

$$\frac{\|d_k\|^2}{(g_k^Td_k)^2}\leqslant\frac{k}{\nabla^2},$$

从而有 $\sum_{k\geqslant 1}\frac{(g_k^Td_k)^2}{\|d_k\|^2}=\infty$，这与引理2矛盾. 定理得证.

参考文献:

[1] Fletcher R, Reeves C. Function minimization by conjugate gradients[J]. Comput J, 1963, 7: 163-168.

[2] Polak E, Ribiere G. Note sur la convergence de directions conjugates [J]. Rev Francaise Informat Recherche Opertionelle 3e Annee, 1969, 16: 35-43.

[3] Hestenes M R, Stiefel E L. Methods of conjugate gradients for soloing linear systems [J]. J Res Nat Bur Standards Sect, 1952, 5(49): 409-436.

[4] Fletcher R. Practical Methods of Optimization[M] (2nd). New York: Wiley-Interscience, 1987. 63-76.

[5] Dai Y H, Yuan Y X. A nonlinear conjugate gradient method with a strong global convergence property [J]. SIAM Journal on Optimization, 1999, 10(1): 177-182.

[6] 陈元媛, 曹兴涛, 杜守强. 一种新的非线性共轭梯度法的全局收敛性[J]. 青岛大学学报, 2004, 17(2): 22-24.

[7] 杜守强, 陈元媛. 一类在新的线搜索下的共轭梯度法[J]. 滨州师专学报, 2002, 18(4): 16-18.

[8] Dai Y H. Conjugate gradient methods with Armijo-type line search [J]. Acta Mathematical Applicate Sinica (English Series), 2002, 18(1): 123-130.

(责任编辑: 黎贞崇)