广西科学 Guangxi Sciences 2006, 13(4):279～281

一种新共轭梯度法的全局收敛性*

A New Conjugate Gradient Method with a Global Convergent Property

田亚娟,马昌凤

TIAN Ya-juan,MA Chang-feng

(桂林电子科技大学计算科学与数学系,广西桂林 541004)

(Deptartment of Computing Science and Mathematics,Guilin University of Electronic Technology, Guilin,Guangxi,541004,China)

摘要:在文献[4,5]的基础上,提出求解无约束优化问题的共轭梯度公式中 β_k 参数的一种新的计算公式:$\beta_{k+1}=\mu\frac{\|g_{k+1}\|^2}{d_k^T y^k}$,$0<\mu<1$;对标准 Wolf 搜索条件进行推广,得到一种新的算法,并证明了算法的全局收敛性.

关键词:无约束优化 共轭梯度法 线搜索 全局收敛性

中图法分类号:O241.7 **文献标识码**:A **文章编号**:1005-9164(2006)04-0279-03

Abstract: We propose a new formula about β_k based on the paper [4] and [5], that is $\beta_{k+1}=\mu\frac{\|g_{k+1}\|^2}{d_k^T y^k}$, $0<\mu<1$. This paper presents a wider line search than the standard Wolf search and give a new conjugate gradient method with global convergence.

Key words: unconstrained optimization, conjugate gradient method, line search, global convergence

考虑无约束优化问题:

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

其中 $f:R^n\rightarrow R^1$ 连续可微. 共轭梯度法是求解问题(1)的一类重要方法[1~3],其迭代公式为:

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

$$d_k=\begin{cases}-g_k, & k=1;\\ -g_k+\beta_k d_{k-1}, & k>1.\end{cases} \tag{3}$$

其中 $g_k=\nabla f(x_k)$,d_k 为搜索方向,α_k 为步长因子,β_k 为参数. β_k 可由以下著名公式给出:

$$\beta_{k+1}=\frac{g_{k+1}^T g_{k+1}}{d_k^T g_k} \quad \text{(Dixon 公式)},$$

$$\beta_{k+1}=\frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \quad \text{(Flectcher-Reeves 公式)}, \tag{4}$$

在文献[4,5]的基础上,本文提出一种新的 β_k 的计算公式:

$$\beta_{k+1}=\mu\frac{\|g_{k+1}\|^2}{d_k^T y^k}, 0<\mu<1, \tag{5}$$

其中 $y^k=g_{k+1}-g_k$,且在文献[6]的基础上对标准 Wolf 搜索条件进行推广,在一定条件下证明算法的全局收敛性.

1 推广的线搜索和新算法

我们知道,步长因子 α_k 的选取应满足一定的下降条件. 设 d_k 为下降方向,即满足 $g_k^T d_k<0$. 精确线性搜索要求步长因子 α_k 满足正交条件:

$$g(x_k+\alpha_k d_k)^T d_k=0;$$

非精确线性搜索的 Wolf 搜索要求步长因子 α_k 满足条件:

$$f(x_k+\alpha_k d_k)-f(x_k)\leqslant\lambda\alpha_k g_k^T d_k, \tag{6}$$

$$g(x_k+\alpha_k d_k)^T d_k>\sigma g_k^T d_k, 0<\lambda<\sigma<1; \tag{7}$$

而强 Wolf 搜索要求步长因子 α_k 满足式(6)和

$$|g(x_k+\alpha_k d_k)^T d_k|\leqslant-\sigma g_k^T d_k, 0<\lambda<\sigma<1;$$

若步长因子 α_k 满足式(6)和搜索条件:

$$\sigma_1 g_k^T d_k\leqslant g(x_k+\alpha_k d_k)^T d_k\leqslant-\sigma_2 g_k^T d_k, \tag{8}$$

其中 $0<\lambda<\sigma_1<1$,σ_2 是任意非负数,则称为放宽了的强 Wolf 搜索.

我们将式(6),(7)换成式(8)和下式:

$$f(x_k+\alpha_k d_k)-f(x_k)\leqslant-\gamma\alpha_k^2\|d_k\|^2, 0<\gamma<1, \tag{9}$$

其中 $0<\gamma\leqslant\sigma_1<1$,$0<\sigma_2<+\infty$. 与式(6)相比,

收稿日期:2005-12-26

修回日期:2006-06-15

作者简介:田亚娟(1980-),女,陕西宝鸡人,硕士研究生,主要从事互补问题研究工作。

*广西自然科学基金项目(0448075)资助。

当$\alpha_k \| d_k \|$比较大时,式(9)能保证使$f(x)$下降的更大.所以推广的搜索条件比标准 Wolf 搜索条件好.

基于上述讨论,我们提出新算法.

算法 1

步骤1　取初始点$x_1 \in R^n$,$k=1$时,$d_1=-g_1$,若$g_1=0$,则停止;

步骤2　计算$\alpha_k>0$满足式(8),(9);

步骤3　计算$g_{k+1}=\nabla f(x_k+\alpha_k d_k)$,若$g_{k+1}=0$,则停;其中$x_{k+1}=x_k+\alpha_k d_k$;

步骤4　计算$\beta_k=\mu\frac{\| g_{k+1} \|^2}{d_k^T y^k}$,$0<\mu<1$,由式(3)得$d_{k+1}$,置$k:=k+1$,转步骤2.

2　新算法的全局收敛性

我们提出2个基本假设:

① f在水平集$\Omega=\{x \in R^n | f(x) \leqslant f(x_1)\}$上有下界;

② $g(x)$是 Lipschitz 连续的,即存在$L>0$,使$\| g(x)-g(y) \| \leqslant L \| x-y \|, \forall\ x,y \in R^n$.

引理 1　若满足假设条件①,②,则在上述新的搜索条件在式(8)和(9)下α_k总是存在的.

证明　设$\omega(\alpha)=f(x_k+\alpha_k d_k)+\gamma\alpha_k^2 \| d_k \|^2$在水平集$\Omega$上有下界,即在整个空间有下界:

$$\lim_{u \to 0^+} \frac{\omega(u)-\omega(0)}{u}=g_k^T d_k<0,$$

则存在$\lambda_1>0$,使$\lambda \in (0,1]$,$\frac{\omega(\alpha)-\omega(0)}{\alpha}<0$.假设

$$\lim_{\alpha \to +\infty} \frac{\omega(\alpha)-\omega(0)}{\alpha}=+\infty,$$

存在$\bar{\alpha}_k$,使得

$$\frac{\omega(\bar{\alpha}_k)-\omega(0)}{\bar{\alpha}_k}=0, \tag{10}$$

且对$\forall\ \alpha \in (0,\bar{\alpha}_k]$都有

$$\frac{\omega(\alpha)-\omega(0)}{\alpha} \leqslant 0. \tag{11}$$

由式(10)利用中值定理得$\omega'(\theta_k \bar{\alpha}_k)=0$及

$$\omega'(\theta_k \bar{\alpha}_k)=g(x_k+\theta_k \bar{\alpha}_k d_k)^T d_k+2\gamma\theta_k \bar{\alpha}_k \| d_k \|^2=0,\quad (0<\theta_k<1),$$

所以

$$g(x_k+\theta_k \bar{\alpha}_k d_k)^T d_k=-2\gamma\theta_k \bar{\alpha}_k \| d_k \|^2 \geqslant -2\sigma_1\alpha_k \| d_k \|^2<0. \tag{12}$$

令$\alpha_k=\theta_k \bar{\alpha}_k$,由式(12)易得$\alpha_k$满足式(8),由式(11)知,$\alpha_k$满足式(9).证毕.

引理 2　若假设条件①,②成立,α_k满足式(8),(9),则$g_k^T d_k<0$对所有k成立.

证明　当$k=1$时,由式(3)得$d_1=-g_1$,故$g_1^T d_1=-g_1^T g_1=-\| g_1 \|^2<0$;假设在第$k$步($k>1$),$g_k^T d_k<0$成立,下面证明该不等式对$k+1$也成立,事实上,

$$g_{k+1}^T d_{k+1}=-\| g_{k+1} \|^2+\beta_k g_{k+1}^T d_k=-\| g_{k+1} \|^2+\mu\frac{\| g_{k+1} \|^2}{\delta_k} g_{k+1}^T d_k=-\frac{\| g_{k+1} \|^2}{\delta_k}(\delta_k-\mu g_{k+1}^T d_k)=-\frac{\| g_{k+1} \|^2}{\delta_k}(d_k^T g_{k+1}-d_k^T g_k-\mu g_{k+1}^T d_k)=\frac{\| g_{k+1} \|^2}{\delta_k} g_k^T d_k. \tag{13}$$

由

$$d_k^T g_k<0, \sigma_1 g_k^T d_k \leqslant g(x_k+\alpha_k d_k)^T d_k \leqslant -\sigma_2 g_k^T d_k,$$

故

$$0<(\sigma_1-1) g_k^T d_k \leqslant \delta_k \leqslant -(1+\sigma_2) g_k^T d_k,$$

$$0<-\frac{1}{(1+\sigma_2) g_k^T d_k} \leqslant \frac{1}{\delta_k} \leqslant \frac{1}{(\sigma_1-1) g_k^T d_k}. \tag{14}$$

上述不等式同乘以$\| g_{k+1} \|^2 g_k^T d_k$得

$$\frac{1}{\sigma_1-1} \| g_{k+1} \|^2 \leqslant \frac{\| g_{k+1} \|^2 g_k^T d_k}{\delta_k} \leqslant -\frac{1}{1+\sigma_2} \| g_{k+1} \|^2<0. \tag{15}$$

故$g_{k+1}^T d_{k+1}<0$.证毕.

根据引理2,算法1所给的搜索方向总是$f(x)$在x_k处的下降方向.

引理 3　若假设条件①,②成立,α_k满足式(8)和(9),则

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

证明　由引理2得$g_k^T d_k<0$,所以$\{f(x_k)\}$为单调下降有下界数列,即$\{f(x_k)\}$为收敛数列.由式(9)得

$$0<(\sigma_1-1) g_k^T d_k \leqslant d_k^T(g_{k+1}-g_k) \leqslant -(1+\sigma_2) g_k^T d_k. \tag{16}$$

利用假设条件②得

$$d_k^T(g_{k+1}-g_k) \leqslant L\alpha_k \| d_k \|^2. \tag{17}$$

故由式(16),(17)知

$$\alpha_k \geqslant \frac{(\sigma_1-1) g_k^T d_k}{L \| d_k \|^2}.$$

再利用式(9)得

$$f_k-f_{k+1} \geqslant \gamma\alpha_k^2 \| d_k \|^2=\frac{\gamma(\sigma_1-1)^2}{L^2} \frac{(g_k^T d_k)^2}{\| d_k \|^2}=c\frac{(g_k^T d_k)^2}{\| d_k \|^2} \forall\ k \in \{1,2,\cdots\}, \tag{18}$$

其中$c=\frac{\gamma(\sigma_1-1)^2}{L^2}$.再由$f(x_k)$的收敛性得

$$\sum_{k \geqslant 1} \frac{(g_k^T d_k)^2}{\| d_k \|^2}<\infty. \tag{19}$$

定理　设$\{x_k\}_{k \geqslant 1}$是由算法1产生的序列,且假设条件①,②成立,则算法或终止于稳定点,或

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

证明　考虑无限序列的情形.用反证法.若结论

不成立，则存在 $\rho > 0$，使

$$\| g_k \| \geqslant \rho, \forall k. \quad (20)$$

由引理 2 知

$$g_{k+1}^T d_{k+1} = \frac{\| g_{k+1} \|^2 g_k^T d_k}{\delta_k} = \frac{1}{\mu} \beta_k g_k^T d_k, \quad (21)$$

则

$$\beta_k = \mu \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \quad (22)$$

将式(3) 写成

$$d_{k+1} + g_{k+1} = \beta_k d_k,$$

故

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

上式两端同除以 $(g_{k+1}^T d_{k+1})^2$，利用式(21) 得

$$\frac{\| d_{k+1} \|^2}{(g_{k+1}^T d_{k+1})^2} = \mu^2 \frac{(g_{k+1}^T d_{k+1})^2 \| d_k \|^2}{(g_k^T d_k)^2 (g_{k+1}^T d_{k+1})^2} - \frac{2}{g_{k+1}^T d_{k+1}} - \frac{\| g_{k+1} \|^2}{(g_{k+1}^T d_{k+1})^2} = \mu^2 \frac{\| d_k \|^2}{(g_k^T d_k)^2} - \left(\frac{1}{\| g_{k+1} \|} + \frac{\| g_{k+1} \|}{g_{k+1}^T d_{k+1}} \right)^2 + \frac{1}{\| g_{k+1} \|^2} \leqslant \mu^2 \frac{\| d_k \|^2}{(g_k^T d_k)^2} + \frac{1}{\| g_{k+1} \|^2}.$$

又因

$$\frac{\| d_1 \|^2}{(g_1^T d_1)^2} = \frac{1}{\| g_1 \|^2}, \| g_k \| \geqslant \rho, 0 < \mu < 1,$$

故

$$\frac{\| d_k \|^2}{(g_k^T d_k)^2} \leqslant \sum_{i=0}^{k-1} \mu^{2i} \frac{1}{\| g_{k+i} \|^2} \leqslant \frac{1}{\rho^2} \sum_{i=0}^{k-1} \mu^{2i} \approx \frac{1}{\rho^2} \frac{1}{1-\mu^2}.$$

上式表明

$$\sum_{k \geqslant 1} \frac{(g_k^T d_k)^2}{\| d_k \|^2} = \infty, \quad (23)$$

这与引理 2 矛盾，故定理成立. 证毕.

参考文献：

［1］ SHI ZHEN JUN. A class of descent methods and its global convergence［J］. Journal of Qufu Normal University, 2000, 26(3): 4-6.

［2］ DU XUE WU. Global convergence of a class of unconstrained minimization methods including the FR method［J］. OR Transactions, 2004, 8(4): 1-9.

［3］ 袁亚湘，孙文瑜. 最优化理论与方法［M］. 北京：科学出版社，1997：98-106.

［4］ 时贞军. 基于共轭梯度法的下降算法［J］. 曲阜师范大学学报，1999，25(1)：5-6.

［5］ 颜世建. 共轭下降法［J］. 南京师大学报，1996，19(2)：12-14.

［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.

（责任编辑：邓大玉　凌汉恩）

（上接第 278 页 Continue from page 278）

所以，文献［14］提出的杂交共轭梯度公式是本文所给一类新的 DY-型共轭梯度公式的特例.

参考文献：

［1］ CHEN Y, CAO X, DU S. A nonlinear conjugate gradient methods with a global convergence property［J］. Jounal of Qingdao University: Natural Science Edition, 2004, 17(2): 22-24.

［2］ HESTENESAND M R, STIEFEL E. Method of conjugate gradient for solving linear equations［J］. J Res Nat Bur Stand, 1952, 49: 409-436.

［3］ FLETCHER R, REEVES C. Function minimization by conjugate gradients［J］. Compute J, 1964, 7: 149-154.

［4］ POLAK E, RIBIERE G. Note Sur la convergence de directions conjugees［J］. Rev Francaise Informat Recherche Operationelle, 3e Année, 1969, 16: 35-43.

［5］ POLYAK B T. The conjugate gradient method in extreme problems［J］. USSR Comp Math and Math Phys, 1969, 9: 94-112.

［6］ FLETCHER R. Practical method of optimization, vol Ⅰ: unconstrained optimization［M］. 2nd edition. New York: Wiley, 1987.

［7］ LIU Y, STOREY C. Efficient generalized conjugate gradient algorithms, part 1: theory［J］. Journal of Optimization Theory and Alication, 1991, 69: 129-137.

［8］ DAI Y, YUAN Y. A nonlinear conjugate gradient with a strong global convergence properties［J］. SIAM Journal of Optimization, 1999, 10: 177-182.

［9］ POWELL M J D. Nonconvex minimization calculations and the conjugate gradient methds, in: Lecture notes in Mathematics vol J［M］. Berlin: Springer-Verlag, 1984: 122-141.

［10］ M Al-BAALI. Descent property and global convergence of the Fletcher-Reeves method with inexact line search［J］. IMA J Numer Anal, 1985, 5: 121-124.

［11］ TOUATI-AHAMED D, STOREY C. Globally convergence hybrid conjugate gradient methods［J］. Joueal of Optimization Theory and Applications, 1990, 64(2): 379-397.

［12］ HAN JIYE, LIU GUANGHUI, YIN HONGXIA. Convergence properties of conjugate gradient methods with strong Wolfe line search［J］. J Systems Science and Mathematics Science, 1998, 11(2): 112-116.

［13］ GILBERT J C, NOCEDAL J. Global convergence of conjugate gradient methods for optimization［J］. SIAM J Optimization, 1992, 2: 21-42.

［14］ DAI Y H, YUAN Y. An efficient hybrid conjugate gradient method for unconstrained optimization［J］. Annals of OR, 2001, 103: 33-47.

（责任编辑：韦廷宗　邓大玉）