DOI: 10.13656/j.cnki.gxkx.2008.04.019
广西科学 Guangxi Sciences 2008, 15(4): 383~ 385

改进的 FR共轭梯度算法及其全局收敛性

Modified FR Conjugate Gradient Method and Its Global Convergence

刘金魁,王开荣*,郑 丽
LIU Jin-kui, WANG Kai-rong, ZHENG Li

(重庆大学数理学院,重庆 400030)
(College of Mathematics and Physics, Chongqing University, Chongqing, 400030, China)

摘要: 给出一种求解无约束优化问题的改进的 FR共轭梯度算法,证明该算法在强 Wolfe线搜索下具有充分下降性和较好的全局收敛性,并用数值试验说明新算法是有效的。

关键词: 无约束优化 共轭梯度法 Wolfe线搜索 充分下降性 全局收敛性
中图法分类号: O224 文献标识码: A 文章编号: 1005-9164(2008)04-0383-03

Abstract A modified FR conjugate gradient method is proposed to solve unconstrained optimization problems. Under the strong Wolfe line search, we proved the sufficient descent property and the preferable global convergence of the modified FR method. Many numerical experiments show that the new method is very efficient.

Key words unconstrained optimization, conjugate gradient method, Wolfe line search, sufficient descent property, global convergence

考虑无约束优化问题: $\min_{x\in R^n} f(x)$,其中 $f: R^n \to R$ 的连续可微函数.求此类问题最重要的方法之一是共轭梯度算法,它的主要迭代格式为:

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

$$d_k = \begin{cases} -g_1, & k = 1, \\ -g_k + \beta_k d_{k-1}, & k \geqslant 2, \end{cases} \tag{2}$$

其中 $g_k = \nabla f(x_k)$, $\alpha_k \geqslant 0$是步长因子,由某种线搜索确定; d_k 是搜索方向, β_k 为标量.

著名的 β_k 公式[1~3]有: $\beta_k^{FR} = \dfrac{\|g_k\|^2}{\|g_{k-1}\|^2}$,

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}.$$

对 FR方法[1]现在已有许多结果,其中 Powell在文献 [4]中证明了 FR方法在精确线搜索下具有全局收敛性.在文献 [5]中 Al-Baali证明了 FR方法在强 Wolfe线搜索下且 $\sigma\in(0,1/2]$具有全局收敛性,但是这种方法的数值结果较差. PRP方法[2,3]具有很好的数值结果但是其全局收敛性较差.为了改善 FR方法的数值效果而保持其好的收敛性,我们在 FR方法的基础上给出一个改进的共轭梯度方法(其中常数 $u > 0$),即

$$\beta_k = \begin{cases} \max\{0, \beta_k^{FR} + \min\{0, -g_k^T g_{k-1}/\|g_{k-1}\|^2\}\}, \\ \quad \text{if} \|g_{k-1}\|^2 \geqslant u\|g_k\| \cdot \|d_{k-1}\|, \\ 0, \text{else}. \end{cases} \tag{3}$$

并在强 Wolfe线搜索条件下证明此方法能够保证搜索方向的充分下降性和全局收敛性. 此方法的数值效果优于 FR方法的数值效果。

1 预备知识及算法

为了证明的需要,对目标函数 $f(x)$ 做假设 (H):

(H_1)目标函数 $f(x)$ 在水平集 $\Omega = \{x | f(x) \leqslant f(x_1)\}$ 有下界,其中 x_1 为初始点;

(H_2)设 $V \subset \Omega$, 目标函数 $f(x)$ 在 V内连续可微,且梯度函数 $g(x)$ 在 V内满足 Lipschitz条件,即存在常数 $L > 0$,使得 $\|g(x) - g(y)\| \leqslant L\|x - y\|$,

---

收稿日期: 2008-02-27
修回日期: 2008-06-25
作者简介: 刘金魁(1982-),男,硕士研究生,主要从事经济系统优化决策与控制研究工作。
* 通讯作者。

$\forall x, y\in V$.

新算法的步骤如下.

步骤1 给定 $x_1\in R^n$, $\varepsilon\geq 0$. 令 $d_1=-g_1$, $k:=1$, 如果 $g_1=0$, 停止.

步骤2 步长 α_k 满足强 Wolfe线搜索:

$$f(x_k+\alpha_k d_k)\leq f(x_k)+\delta\alpha_k g_k^T d_k; \qquad (4)$$

$$|g(x_k+\alpha_k d_k)^T d_k|\leq \sigma g_k^T d_k. \qquad (5)$$

其中 $0<\delta<\sigma<1/2$.

步骤3 $x_{k+1}=x_k+\alpha_k d_k$, $g_{k+1}=g(x_{k+1})$, 如果 $\|g_{k+1}\|\leq\varepsilon$, 停止.

步骤4 由公式(3)计算 β_{k+1},由(2)式求 d_{k+1}.

步骤5 设 $k:=k+1$, 转步骤2.

2 算法的全局收敛性

引理1[6] 设目标函数 $f(x)$ 满足假设(H),考虑一般方法 $x_{k+1}=x_k+\alpha_k d_k$,其中 d_k 满足 $g_k^T d_k<0$, 步长 α_k 由强 Wolfe线搜索(4)、(5)求得,则对于任意 $k\in N$, Zoutendijk条件成立,即

$$\sum_{k\geq 1}(g_k^T d_k)^2/\|d_k\|^2<+\infty. \qquad (6)$$

引理2 考虑一般方法(1)、(2),步长 α_k 满足强 Wolfe线搜索(4)、(5),当 β_k 取(3)时,则对任意的 $k>0$,有

$$\frac{1-2\sigma+\sigma^k}{1-\sigma}\leq\frac{-g_k^T d_k}{\|g_k\|^2}\leq\frac{1-\sigma^k}{1-\sigma}. \qquad (7)$$

证明 对 $k=1$,因为 $d_1=-g_1$,(7)式显然成立。现在假设对任意 $k-1$,(7)式成立,将(2)式两端与 g_k 作内积,整理可得

$$-g_k^T d_k/\|g_k\|^2=1-\beta_k g_k^T d_{k-1}/\|g_k\|^2 \qquad (8)$$

由(3)式可知, $0\leq\beta_k\leq\beta_k^{FR}$. (9)

利用(5)式、(8)式、(9)式可得,

$$1+\sigma g_{k-1}^T d_{k-1}/\|g_{k-1}\|^2\leq -g_k^T d_k/\|g_k\|^2\leq 1-\sigma g_{k-1}^T d_{k-1}/\|g_{k-1}\|^2. \qquad (10)$$

于是,利用归纳假设及(10)式的第1个不等式,得

$$-g_k^T d_k/\|g_k\|^2\geq 1-\sigma\frac{1-\sigma^{k-1}}{1-\sigma}=\frac{1-2\sigma+\sigma^k}{1-\sigma}.$$

同理,得

$$-g_k^T d_k/\|g_k\|^2\leq 1+\sigma\frac{1-\sigma^{k-1}}{1-\sigma}=\frac{1-\sigma^k}{1-\sigma}.$$

故(7)式对 k 也成立.由归纳法知,引理成立.

定理1 设目标函数 $f(x)$ 满足假设(H),考虑(1)式、(2)式,其中 β_k 取(3)式,步长 α_k 满足强 Wolfe线搜索(4)、(5),则对新算法产生点列 $\{x_k\}$ 或者 $g_k=0$ 对某个 k 成立,或者

$$\lim_{k\to\infty}\|g_k\|=0. \qquad (11)$$

证明 由(7)式左端可知, $-g_k^T d_k\geq c\|g_k\|^2$,其中 $c=(1-2\sigma)/(1-\sigma)$. 于是,由(6)式得 $\sum_{k\geq 1}c^2\|g_k\|^4/\|d_k\|^2<+\infty$,即

$$\sum_{k\geq 1}\|g_k\|^4/\|d_k\|^2<+\infty. \qquad (12)$$

当 $\|g_{k-1}\|^2<u\|g_k\|\cdot\|d_{k-1}\|$ 时,显然有 $\beta_k=0$,于是,根据(2)式可得 $\|d_k\|=\|g_k\|$.否则,对(2)式两端取模平方,整理得

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

显然,可得

$$\|d_k\|^2\leq\|g_k\|^2+\beta_k^2\|d_{k-1}\|^2+2|\beta_k|\cdot|g_k^T d_{k-1}|.$$

利用(5)式及(7)式右端,得

$$\|d_k\|^2\leq\|g_k\|^2+\beta_k^2\|d_{k-1}\|^2+\frac{2\sigma(1-\sigma^{k-1})|\beta_k|\cdot\|g_{k-1}\|^2}{(1-\sigma)}.$$

再根据(9)式及 $\|g_{k-1}\|^2\geq u\|g_k\|\cdot\|d_{k-1}\|$ 可得

$$\|d_k\|^2\leq\|g_k\|^2+\frac{\|g_k\|^4}{(u\|g_k\|\cdot\|d_{k-1}\|)^2}\|d_{k-1}\|^2+\frac{2\sigma(1-\sigma^{k-1})\|g_k\|^2}{(1-\sigma)}\leq(1+1/u^2+2\sigma(1-\sigma^{k-1})/(1-\sigma))\|g_k\|^2\leq(1/u^2+(1+\sigma)/(1-\sigma))\|g_k\|^2.$$

取 $t^2=1/u^2+(1+\sigma)/(1-\sigma)$,则 $\|d_k\|/\|g_k\|\leq t$.

综上所述, $\|d_k\|/\|g_k\|$ 有界。

令 $r=\max\{t,1\}$,于是,对任意的 k,有

$$\|d_k\|/\|g_k\|\leq r. \qquad (13)$$

根据(12)及(13)式得 $\sum_{k\geq 1}\|g_k\|^2<+\infty$, 所以有 $\lim_{k\to\infty}\|g_k\|=0$ 成立。

3 算法的数值实验

利用 MATLAB编制程序,在新算法中取 β_k 为(3)式,参数 $\delta=0.01$, $\sigma=0.1$, $u=0.005$ or 0.25. 终止条件为: $\|g_k\|\leq 1.0\times 10^{-6}$,或者最大迭代次数大于9999,在强 Wolfe线搜索条件下对文献[7]中的测试函数进行试验,并与 FR方法[7]的数值结果做了比较,详见表1.表1结果表明本文算法数值效果较好.

表1中的 Problem表示文献[7]中测试函数的名称; Dim表示测试函数的维数;数据表示为 NI/NF/NG,它们依次为迭代次数、函数值计算次数、梯度值计算次数。VFR表示本文新算法.

表 1 数值实验结果

Table 1 Numerical results

Problem	Dim	NI/NF/NG		
		FR	VFR ($u=$ 0.005)	VFR ($u=$ 0.005)
Rosenbrock	2	119/349/302	26/136/111	43/200/166
Biggs EXP6	6	255/731/639	212/626/549	226/929/824
Beale	2	49/149/124	67/175/142	67/175/142
Helical valley	3	39/116/97	30/97/82	45/157/131
Bard	3	28/98/81	18/67/54	36/143/123
Wood	4	103/307/251	38/159/128	64/271/225
Kowalik and Osborne	4	372/1066/931	94/293/256	80/302/265
Brown and Dennis	4	55/191/149	69/193/165	69/193/165
Extended Rosenbrock	500	126/371/320	29/149/123	46/214/178
	1000	132/390/337	29/149/123	46/214/178
Penalty I	50	1723/3169/3092	81/402/331	89/573/465
	100	41/180/141	46/238/188	33/248/196
Trigonometric	100	322/460/459	56/125/118	53/121/113
	200	336/468/467	61/130/125	60/131/122
Discrete integral equation	500	7/15/8	6/13/7	6/13/7
	1000	7/15/8	6/13/7	6/13/7
Broyden tridiagonal	500	48/104/99	35/78/73	35/78/73
	1000	65/139/135	35/79/75	35/79/75

参考文献:

[1] Fletcher R, Reeves C. Function minimization by conjugate gradients [J]. Computer Journal, 1964, 7: 149-154.

[2] Polak E, Ribire G. Note sur la xonvergence de directions conjugees [J]. Rev Francaise Informat Recherche Operatinelle 3e Annee, 1969, 16: 35-43.

[3] Polak B T. The conjugate gradient method in extreme problems [J]. USSR Comput Math Math Phys, 1969, 9: 94-112.

[4] Powell M J D. Nonconvex minimization calculation and the conjugate gradient method [M] // Lecture notes in Mathematics vol 1006. Berlin: Springer-Veriag, 1984: 122-141.

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

[6] Zoutendijk G. Nonlinear programming computational methods [M] // Abedie J Integer and Nonlinear Programming North Holland, Amsterdam, 1970: 37-86.

[7] More J J, Garbow B S, Hillstrome K E. Testing unconstrained optimization software [J]. ACM Trains Math Software, 1981, 7: 17-41.

(责任编辑: 邓大玉)

我国实现芯片玻色——爱因斯坦凝聚体

随着物理科学技术的发展,超冷原子介质在超高精度原子频率标准、原子干涉仪、量子信息存储和信息处理等方面获得了重要应用。但是获得超冷原子气体和原子芯片上的玻色——爱因斯坦凝聚体(BEC)的实验装置过于复杂和庞大,而且价格十分昂贵,在一定程度上阻碍了其向应用技术的发展。因此,研制小型化的冷原子实验装置(即原子芯片实验装置)成为国际上冷原子应用技术研究的重要发展方向。原子芯片是利用成熟的半导体制作工艺和MEMS技术,将磁场系统和光学系统集成到一块硅基底芯片上。利用表面电流产生的近表面梯度磁场形成芯片原子磁阱,进而实现集成化的冷原子实验装置。芯片磁阱对原子的束缚非常紧,蒸发冷却时可以快速实现BEC相变,原子芯片不仅提高了冷原子装置的稳定性、可靠性和便携性,而且能够实现一些宏观冷原子装置所不能实现的功能。

国际上只有美国、法国、德国、日本等少数发达国家拥有芯片BEC。我国最近由中国科学院院士、中科院上海光机所量子光学重点实验室王育竹领导的研究人员研究实现了我国第一个原子芯片上的BEC。实现芯片BEC是我国冷原子研究和量子信息存储技术研究的重大标志性进展。

(据《科学时报》)