运筹学与控制论

DOI:CNKI:50-1165/N. 20120915. 2319. 004

一类修正 LS 谱共轭梯度法的全局收敛性*

胡　鹏[1,3]，杜学武[1]，郭翠峰[2]

(1.重庆师范大学 数学学院，重庆 401331；2.西华大学 数学与计算机科学学院，成都 610039；
3.泸州高级中学，四川 泸州 646000)

摘要:谱共轭梯度法是一类将共轭梯度法和谱梯度法相结合的方法。2001 年由 Birgin 和 Martinez 首先提出,但该方法不能保证始终产生下降方向。本文用已有的修正方法,给出一个修正的 Liu-Storey 公式,并结合谱梯度法,提出了一个具有充分下降性的修正 Liu-Storey 谱共轭梯度法,证明了该方法在标准 Armijo 非精确线搜索下的全局收敛性,并易推知该方法在 Armijo-Goldstein 非精确线搜索准则下同样满足全局收敛性。给出的数值实验表明,新算法略优于 LS 方法。

关键词:修正的 Liu-Storey 共轭梯度法；Armijo 型线搜索；全局收敛性

中图分类号:O221.2　　**文献标志码**:A　　**文章编号**:1672-6693(2012)05-0013-03

首先考虑下面的无约束最小化问题:

$$\min \quad f(x),x\in \mathbf{R}^n \tag{1}$$

其中 $\mathbf{R}^n$ 表示维欧氏空间,$f:\mathbf{R}^n\to\mathbf{R}$ 是连续可微函数,并且它的梯度 g 是可获得的。共轭梯度法因其具有算法简单,易于编程,需要存储空间小等优点而被广泛应用于大规模无约束问题(1) 式。它的迭代格式

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

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

其中 d_k 为搜索方向,α_k 是通过一维搜索获得的步长。最早的非线性共轭梯度法是由 Fletcher 和 Reeves[1] 在 1964 年提出的,简称 FR 方法。FR 方法的参数 β_k 的公式为 $\beta_k^{FR}=\dfrac{\|g_k\|^2}{\|g_{k-1}\|^2}$。比较著名的共轭梯度法还有 PRP、HS、LS 和 DY 方法[2-6]。

在共轭梯度法的计算中,还需考虑确定步长 α_k 的线搜索准则,包括精确的和非精确的。常用的非精确线搜索有 Wolfe 线搜索[7]、强 Wolfe 线搜索、Armijo 线搜索、Armijo-Goldstein 线搜索等。Armijo 线搜索准则为

$$f(x_k+\alpha_k d_k)\leqslant f(x_k)+\sigma\alpha_k g_k^{\mathrm T}d_k \tag{4}$$

选择公式中步长 $\alpha_k=\max\{\rho^j, j=0,1,2,\cdots\}$,其中 $\rho\in(0,1)$, $\sigma\in(0,1)$。

Armijo-Goldstein 线搜索准则

$$f(x_k)+(1-\sigma)\alpha_k g_k^{\mathrm T}d_k\leqslant f(x_k+\alpha_k d_k)\leqslant f(x_k)+\sigma\alpha_k g_k^{\mathrm T}d_k \tag{5}$$

式中步长 $\alpha_k=\max\{\rho^j, j=0,1,2,\cdots\}$,其中 $\rho\in(0,1)$, $\sigma\in\left(0,\dfrac{1}{2}\right)$。

Birgin and Martinez[8] 结合共轭梯度法和谱梯度法[9] 构造了一种谱共轭梯度法。搜索方向 d_k 定义为

$$d_k=\begin{cases}-g_k, k=1\\ -\theta_k g_k+\beta_k d_{k-1}, k\geqslant 2\end{cases},\beta_k=\frac{(\theta_k y_{k-1}-S_{k-1})^{\mathrm T}g_k}{d_{k-1}^{\mathrm T}y_{k-1}}$$

其中 $y_{k-1}=g_k-g_{k-1}, s_{k-1}=x_k-x_{k-1}, \theta_k=\dfrac{s_{k-1}^{\mathrm T}s_{k-1}}{s_{k-1}^{\mathrm T}y_{k-1}}$ 是一个参数。但是该方法不能保证搜索方向 d_k 具有充分下降性。韦增欣等在文献[10]中对 HS 方法进行修正,并在一定的线搜索下证明了算法的收敛性,数值实验表明文献[10]中的修正 HS 方法计算效果优于 HS 方法。

由于 LS 方法的参数 β_k 与 HS 方法的参数 β_k 的分子相同,借鉴文献[10]的修正方法,并结合文献[8],本文给出了一类具有充分下降性的修正 LS 谱共轭梯度法,并在标准 Armijo 非精确线搜索下证明了该方法的全局收敛性,给出的数值结果显示该方法略优于 LS 方法。

1 LS 谱共轭梯度算法

考虑无约束优化问题(1)式,结合文献[5,8,10-11],本文给出的修正 LS 谱共轭梯度法的迭代格式及参数公式如下

$$x_{k+1}=x_k+\alpha_k d_k \tag{6}$$

* **收稿日期**:2011-11-02　**修回日期**:2012-03-26　**网络出版时间**:2012-9-15 23:19

资助项目:国家自然科学基金(No. 10971241;No. 11171363);重庆师范大学自然科学基金(No. 08XLR022)

作者简介:胡鹏,男,硕士研究生,研究方向为最优化理论与算法;**通讯作者**:杜学武,E-mail:duxuewu@cqnu.edu.cn

网络出版地址:http://www.cnki.net/kcms/detail/50.1165.N.20120915.2319.201205.13_004.html

$$d_k=\begin{cases}-g_k,k=1\\-\theta_k g_k+\beta_k^{\mathrm{MLS}}d_{k-1},k\geqslant 2\end{cases}\tag{7}$$

其中 β_k^{MLS} 表示修正的 LS 公式,它的形式为

$$\beta_k^{\mathrm{MLS}}=\frac{g_k^{\mathrm{T}}\left(g_k-\frac{g_k^{\mathrm{T}}g_{k-1}}{\|g_{k-1}\|^2}g_{k-1}\right)}{-g_{k-1}^{\mathrm{T}}d_{k-1}}\tag{8}$$

(7)式中 $\theta_k=1-\frac{g_k^{\mathrm{T}}d_{k-1}(1-\cos^2 r_k)}{g_{k-1}^{\mathrm{T}}d_{k-1}}$ 是一个参数,r_k 表示 g_k 和 g_{k-1} 之间的夹角。由(6)、(7)式易得对所有的 k 均有

$$g_k^{\mathrm{T}}d_k=-\|g_k\|^2\tag{9}$$

成立。因此新给出的修正 LS 谱共轭梯度法是一个下降方法。

算法1 Step1 给定 $\rho\in(0,1)$,$\sigma\in(0,1)$,$\varepsilon>0$。选择一个初始点 $x_1\in\mathbf{R}^n$,并令 $d_1:=-g_1$,$k:=1$;

Step2 如果 $\|g_k\|\leqslant\varepsilon$,则停止。反之进入 Step3;

Step3 让 $x_{k+1}=x_k+\alpha_k d_k$,d_k 由(7)式计算得到,要求 α_k 满足 Armijo 线搜索准则(4)式。

Step4 令 $k:=k+1$,并转入 Step2。

2 全局收敛结果

为了证明需要,要求目标函数满足假设 H:ⅰ) $f(x)$ 在水平集 $L=\{x|f(x)\leqslant f(x_1)\}$ 上有下界,其中 x_1 为初始点;ⅱ) 在 L 的一个邻域 N 内,$f(x)$ 连续可微且梯度向量 g 满足 Lipschitz 条件,即存在常数 $M>0$ 使得 $\|g(x)-g(y)\|\leqslant M\|x-y\|$,$\forall x,y\in N$。

从上述假设 H 知,一定存在两个正常数 β 和 r,使得 $\|x\|\leqslant\beta$,$\|g(x)\|\leqslant r$,$\forall x\in N$ 成立。

引理1 若假设 H 成立,$\{g_k\}$ 和 $\{d_k\}$ 由修正的 LS 谱共轭梯度算法产生,则一定存在一个常数 $c=\min\left\{1,\frac{(1-\sigma)\rho}{M}\right\}>0$ 在 $d_k\neq 0$ 的前提下,对所有的 k 使得不等式(10)式成立

$$\alpha_k\geqslant c\frac{\|g_k\|^2}{\|d_k\|^2}\tag{10}$$

证明 当 $\alpha_k=1$,由(9)式和 Schwarz 不等式可得 $\|g_k\|\cdot\|d_k\|\geqslant|g_k^{\mathrm{T}}d_k|=\|g_k\|^2$,变形得 $\frac{\|g_k\|}{\|d_k\|}\leqslant 1$。因此,当 $c=1$ 时,(10)式成立。

若 $\alpha_k<1$,由 Armijo 线搜索准则知,$\rho^{-1}\alpha_k$ 不满足不等式(4)式,即

$$f(x+\rho^{-1}\alpha_k d_k)-f(x_k)>\sigma\rho^{-1}\alpha_k g_k^{\mathrm{T}}d_k\tag{11}$$

成立。因为算法1是一个下降方法,所以有 $\{x_k\}\subset L$。由中值定理和假设 H 中的ⅱ)可知对一些 $t_k\in(0,1)$,有

$$\begin{aligned}&f(x_k+\rho^{-1}\alpha_k d_k)-f(x_k)=\\&\rho^{-1}\alpha_k g(x_k+t_k\rho^{-1}\alpha_k d_k)^{\mathrm{T}}d_k=\rho^{-1}\alpha_k g_k^{\mathrm{T}}d_k+\\&\rho^{-1}\alpha_k(g(x_k+t_k\rho^{-1}\alpha_k d_k)-g_k)^{\mathrm{T}}d_k\leqslant\\&\rho^{-1}\alpha_k g_k^{\mathrm{T}}d_k+M\rho^{-2}\alpha_k^2\|d_k\|^2\end{aligned}\tag{12}$$

结合(11)、(12)和(9)式得

$$\alpha_k\geqslant\frac{(1-\sigma)\rho\|g_k\|^2}{M\|d_k\|^2}\geqslant c\frac{\|g_k\|^2}{\|d_k\|^2}$$ 证毕

引理2 若假设 H 成立,$\{g_k\}$ 和 $\{d_k\}$ 由算法1产生,则

$$\sum_{k\geqslant 1}\frac{\|g_k\|^4}{\|d_k\|^2}<\infty\tag{13}$$

证明 在假设 H 的条件下,结合(4)式得 $\sum_{k\geqslant 0}(-\sigma\alpha_k g_k^{\mathrm{T}}d_k)<\infty$,结合(9)式,即得

$$\sum_{k\geqslant 0}(\alpha_k^2\|d_k\|^2)<\infty\tag{14}$$

综合(10)、(14)式,可证明(13)式成立。 证毕

定理1 若假设 H 成立,$\{x_k\}$ 由算法1产生,此时算法1或者有限步终止于一个平稳点,或者以(15)式的意义收敛

$$\liminf_{k\to\infty}\|g_k\|=0\tag{15}$$

证明 若对某个 k 有 $g_k=0$,则 x_k 为稳定点,定理得证,否则假设(15)式不成立,则存在 $\varepsilon>0$,使得

$$\|g_k\|^2\geqslant\varepsilon,\ \forall k\geqslant 1\tag{16}$$

由(7)式得

$$\|d_k\|^2=(\beta_k^{\mathrm{MLS}})^2\|d_{k-1}\|^2-2\theta_k g_k^{\mathrm{T}}d_k-\theta_k^2\|g_k\|^2\tag{17}$$

在(17)式两边同时除 $\|g_k\|^4$,并结合(8)、(9)和(16)式,可得

$$\begin{aligned}&\frac{\|d_k\|^2}{\|g_k\|^4}=\frac{(\beta_k^{\mathrm{MLS}})^2\|d_{k-1}\|^2}{(g_k^{\mathrm{T}}d_k)^2}-\frac{2\theta_k g_k^{\mathrm{T}}d_k}{(g_k^{\mathrm{T}}d_k)^2}-\frac{\theta_k^2}{\|g_k\|^2}=\\&\left[\frac{\|g_k\|^2\cdot\|g_{k-1}\|^2(1-\cos^2\gamma_k)}{-\|g_{k-1}\|^2(g_{k-1}^{\mathrm{T}}d_{k-1})}\right]^2\cdot\\&\frac{\|d_{k-1}\|^2}{\|g_k\|^2}-\frac{(\theta_k^2-2\theta_k)}{\|g_k\|^2}=\\&(1-\cos^2\gamma_k)^2\frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4}-\frac{(\theta_k-1)^2}{\|g_k\|^2}+\frac{1}{\|g_k\|^2}\leqslant\\&\frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4}+\frac{1}{\|g_k\|^2}\leqslant\frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4}+\frac{1}{\varepsilon}\end{aligned}\tag{18}$$

因为 $\frac{\|d_1\|^2}{\|g_1\|^4}=\frac{1}{\|g_1\|^2}\leqslant\frac{1}{\varepsilon}$,由(18)式可推得

$$\frac{\|d_k\|^2}{\|g_k\|^4}\leqslant\frac{k}{\varepsilon},\ \forall k\geqslant 1\tag{19}$$

由(19)式可知 $\sum_{k\geqslant 1}\frac{\|g_k\|^4}{\|d_k\|^2}=\infty$,这与(13)式相矛盾。

证毕

定理2 若假设 H 成立,将算法1中的 Armijo 线搜索准则改为 Armijo-Goldstein 线搜索准则,$\{x_k\}$ 由修改后的算法产生。新算法或者有限步终止于一个平稳点,或者 $\liminf_{k\to\infty}\|g_k\|=0$。

证明与定理 1 类似。

3 数值试验

本文给出的方法简记为 SMLS,本节将在标准 Armijo 非精确线搜索准则下,分别用 LS 方法、SMLS 方法对测试函数进行试算,并进行比对。表格中 Problem 表示测试函数[12],NI/NF/NG 分别代表迭代次数、函数迭代次数、梯度迭代次数。计算中参数 $\sigma=0.5$,$\rho=0.8$,取 $\varepsilon=10^{-5}$。

表 1 LS 方法与 SMLS 方法的数值结果对比

Fig. 1 The numerical results contrast between LS method and SMLS method

	LS 方法	SMLS 方法
测试函数	NI/NF/NG	NI/NF/NG
CUBE	36/1559/37	72/1995/73
ROSE*	30/823/31	30/438/31
Powell S*	77/1404/78	69/1229/70
Wood*	189/5183/190	176/4543/177
FROTH	9/260/10	12/470/13

表 1 中,加"*"的测试函数表示 SMLS 算法的数值结果优于 LS 算法。从上表给出的数值结果可知,在标准 Armijo 非精确线搜索准则下,本文给出的 MSLS 方法的计算效果略优于 LS 方法。

参考文献:

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

[2] Polak E, Ribiere G. Note sur la convergence de directions conjugees [J]. Rev Francaise Informat Recherche Operatinelle, 1969, 16: 35-43.

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

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

[5] Liu Y, Storey C. Efficient generalized conjugate gradient algorithms [J]. JOTA, 1991, 69(1): 129-152.

[6] Dai Y H, Yuan Y X. A nonlinear conjugate gradient with a strong global convergence property [J]. SIAM Journal on Optimization, 2000, 10(1): 177-182.

[7] Powell M J D. Nonconvex minimization calculations and the conjugate gradient method [M]. Lecture Notes in Mathematics, 1984, 1066: 122-141.

[8] Birgin E G, Martinez J M. A spectral conjugate gradient method for the large scale unconstrained optimization [J]. Appl Math Optim, 2001, 43(2): 117-128.

[9] Raydan M. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem [J]. SIAM J Optim, 1997, 7(1): 26-33.

[10] Wei Z X, Huang H D, Tao Y R. A modified Hestenes-Stiefel conjugate gradient method and its Convergence [J]. Journal of Mathematical Research & Exposition, 2010, 30(2): 297-308.

[11] Cao W, Wang K R, Wang Y L. Global convergence of a modified spectral CD conjugate gradient method [J]. Journal of Mathematical Research & Exposition, 2011, 31(2): 261-268.

[12] Mor'e J J, Garbow B S, Hillstrom K E. Testing unconstrained optimization software [J]. ACM Trans Math Software, 1981, 7(1): 17-41.

Operations Research and Cybernetics

Global Convergence of a Modified LS Spectral Conjugate Gradient Method

HU Peng[1,3], DU Xue-wu[1], GUO Cui-feng[2]

(1. College of Mathematics Science, Chongqing Normal University, Chongqing 401331;

2. College of Mathematics and Computer, Xihua University, Chengdu 610039;

3. Luzhou Senior Middle School, Luzhou Sichuan 646000, China)

Abstract: Spectral conjugate gradient method is a kind of method that combines conjugate gradient method with spectral gradient method. In 2001, it was first put forward by Birgin and Martinez in [8], but this method can not always guarantee to generate descent directions. This paper first gives out a modified Liu-Storey formula which uses the modified method given by literature [10], and then combines the modified Liu-Storey formula with the spectral gradient method, putting forward a modified Liu-Storey spectral conjugate gradient method satisfying the sufficient descent condition. And the global convergence of the method with the standard Armijo inexact line search is proved; it is easy to deduce the method also satisfying the global convergence under the Armijo-Goldstein inexact line search rule. The given numerical results show that the new method is a little better than LS method.

Keywords: modified Liu-Storey conjugate gradient; Arrmijo-type line search; global convergence

(责任编辑 黄 颖)