运筹学与控制论 DOI:10.11721/cqnuj20130403

一类充分下降的谱共轭梯度法*

陈龙卫[1,2]，夏福全[3]，贾朝勇[3]
(1.南京航空航天大学 理学院，南京 210016；2.泰州机电高等职业技术学校 自动化系，江苏 泰州 225300；
3.蚌埠学院 数理系，安徽 蚌埠 233030)

摘要:首先基于共轭梯度法的共轭条件和下降性，提出了一类充分下降的谱共轭梯度法。该方法将经典共轭梯度法中搜索方向由原来的只满足一个共轭条件改变为同时满足一个共轭条件和一个下降条件;然后，在 Wolfe 线搜索下用反证法证明了新算法的全局收敛性;最后,通过 12 个算例，将新算法和已有 SHS 算法在迭代次数和计算时间方面进行了数值比较实验，比较结果表明新算法在这两个方面都明显优越于 SHS 算法。算法的全局收敛性和数值结果的优越性表明，新算法是一个值得研究的方法。

关键词:无约束优化;谱共轭梯度法;充分下降条件;共轭条件;全局收敛
中图分类号:O224 **文献标志码**:A **文章编号**:1672-6693(2013)04-0010-05

考虑无约束优化问题

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

其中 $f:\mathbf{R}^n\to\mathbf{R}$ 为连续可微函数。

共轭梯度法是求解问题(1)的有效方法[1]。经典共轭梯度法的一般迭代格式为

$$x_{k+1}=x_k+\alpha_k d_k,d_k=\begin{cases}-g_0,k=0\\-g_k+\beta_k d_{k-1},k\geqslant 1\end{cases}$$

其中 x_0 为初始迭代点,α_k 是步长因子,d_k 为搜索方向,$g_k=\nabla f(x_k)$,β_k 为参数。β_k 的选取有很多著名的公式,如

$$\beta_k^{\mathrm{HS}}=\frac{g_k^{\mathrm{T}}y_{k-1}}{d_{k-1}^{\mathrm{T}}y_{k-1}},\beta_k^{\mathrm{FR}}=\frac{g_k^{\mathrm{T}}g_k}{g_{k-1}^{\mathrm{T}}g_{k-1}},\beta_k^{\mathrm{DY}}=\frac{g_k^{\mathrm{T}}g_k}{d_{k-1}^{\mathrm{T}}y_{k-1}}$$

其中 $y_{k-1}=g_k-g_{k-1}$。不同的选取公式构成了不同的方法,上述 3 个公式对应的共轭梯度法分别叫做 HS 方法、FR 方法和 DY 方法。其它的参见文献[2]。

2001 年,Birgin 和 Martinez 提出了一种谱共轭梯度方法[3],其搜索方向定义为 $d_k=\begin{cases}-g_0,k=0\\-\theta_k g_k+\beta_k d_{k-1},k\geqslant 1\end{cases}$,其中 $\theta_k=\frac{s_{k-1}^{\mathrm{T}}s_{k-1}}{s_{k-1}^{\mathrm{T}}y_{k-1}}$为谱参数,$\beta_k=\frac{(\theta_k y_{k-1}-s_{k-1})^{\mathrm{T}}g_k}{d_{k-1}^{\mathrm{T}}y_{k-1}}$,$s_{k-1}=x_k-x_{k-1}=\alpha_{k-1}d_{k-1}$。较为遗憾的是,Birgin 和 Martinez 提出的谱共轭梯度法的搜索方向 d_k 不满足下降条件 $d_k^{\mathrm{T}}g_k<0$ 。为了获得全局收敛性，Zhang 等在 FR 方法的基础上构造了 FR 型谱共轭梯度法[4],其搜索方向 d_k 中 $\theta_k=d_{k-1}^{\mathrm{T}}y_{k-1}/\|g_{k-1}\|^2$,$\beta_k=\beta_k^{\mathrm{FR}}$。此方法的一个重要特点是满足充分下降条件

$$d_k^{\mathrm{T}}g_k\leqslant -c\|g_k\|^2 \tag{2}$$

其中 $c>0$ 是常数。另外,他们还证明了此方法在 Armijo 线搜索下的全局收敛性。此后 Du 和 Chen、Lu 等又对文献[4]提出的 FR 型谱共轭梯度法进行了补充和修正[5-6],获得了很好的收敛性和数值结果。王开荣、Cao 等提出了满足充分下降条件(2)的 CD 型谱共轭梯度法[7-8];Du 和 Liu 提出了满足充分下降条件(2)的 HS 型谱共轭梯度法[9];Wan,黄海等提出了满足充分下降条件(2)的 PRP 型谱共轭梯度法[10-11]。

* **收稿日期**:2012-09-06 **修回日期**:2012-11-19 **网络出版时间**:2013-07-20 19:23
作者简介:陈龙卫,男,讲师,硕士,研究方向为数学规划算法,E-mail: tzveclw@sina.com
网络出版地址:http://www.cnki.net/kcms/detail/50.1165.N.20130720.1923.201304.10_029.html

本文在上述文献的基础上，根据充分下降条件(2)及 Dai-Liao 共轭条件[12] $d_k^{\mathrm{T}} y_{k-1}=-t g_k^{\mathrm{T}} s_{k-1}$ ($t \geqslant 0$ 是常数)选取另一组参数，得到了一类充分下降的谱共轭梯度法，并给出算法及其收敛性证明，最后进行数值实验，并分析实验结果。

1 一类充分下降的谱共轭梯度法

考虑一般的谱共轭梯度法

$$d_k=\begin{cases}-g_0, k=0 \\ -\theta_k g_k+\beta_k d_{k-1}, k \geqslant 1\end{cases} \tag{3}$$

其中 θ_k 是谱系数，β_k 是参数。

若取 $\theta_k=\frac{d_{k-1}^{\mathrm{T}} y_{k-1}}{\|g_{k-1}\|^2}$，$\beta_k=\beta_k^{\mathrm{FR}}$，就得到 FR 型谱共轭梯度法方法[4-5]；若取 $\theta_k=1-\frac{g_k^{\mathrm{T}} d_{k-1}}{d_{k-1}^{\mathrm{T}} g_{k-1}}$，$\beta_k=\beta_k^{\mathrm{CD}}$，就得到 CD 型谱共轭梯度法方法[7-8]；若取 $\theta_k=1-\frac{g_k^{\mathrm{T}} g_{k-1} \cdot g_k^{\mathrm{T}} d_{k-1}}{\|g_k\|^2 \cdot d_{k-1}^{\mathrm{T}} y_{k-1}}$，$\beta_k=\beta_k^{\mathrm{HS}}$，就得到 HS 型谱共轭梯度法方法[9]；若取 $\theta_k=\frac{d_{k-1}^{\mathrm{T}} y_{k-1}}{\|g_{k-1}\|^2}-\frac{g_k^{\mathrm{T}} g_{k-1} \cdot g_k^{\mathrm{T}} d_{k-1}}{\|g_k\|^2\|g_{k-1}\|^2}$，$\beta_k=\beta_k^{\mathrm{PRP}}$，就得到 PRP 型谱共轭梯度法方法[10]。

本文将考虑选取另一组参数使其满足

$$d_k^{\mathrm{T}} y_{k-1}=-\theta g_k^{\mathrm{T}} s_{k-1} \tag{4}$$

$$d_k^{\mathrm{T}} g_k=-c g_k^{\mathrm{T}} g_k \tag{5}$$

其中 $0 \leqslant \theta<0.5$，$c>0.5$。$k \geqslant 1$ 时，令

$$\begin{gathered}\Delta_k=d_{k-1}^{\mathrm{T}} y_{k-1}\|g_k\|^2-g_k^{\mathrm{T}} y_{k-1} d_{k-1}^{\mathrm{T}} g_k, \theta_k^1= \\ c\|g_k\|^2 d_{k-1}^{\mathrm{T}} y_{k-1}-\theta d_{k-1}^{\mathrm{T}} g_k g_k^{\mathrm{T}} s_{k-1}, \beta_k^1=\|g_k\|^2\left(c g_k^{\mathrm{T}} y_{k-1}-\theta g_k^{\mathrm{T}} s_{k-1}\right)\end{gathered} \tag{6}$$

若 $\Delta_k \neq 0$，则由(4)、(5)式得

$$\theta_k=\frac{\theta_k^1}{\Delta_k}, \quad \beta_k=\frac{\beta_k^1}{\Delta_k} \tag{7}$$

为了保证较好的收敛性，将新方法的搜索方向修正为

$$d_k=\begin{cases}-g_k, k=0 \text{ 或 } |\cos \eta_k|<\theta' \\ \frac{m_k}{\Delta_k}, k \geqslant 1 \text{ 且 } |\cos \eta_k| \geqslant \theta'\end{cases} \tag{8}$$

其中 $0<\theta'<1$，且 η_k 为 m_k 与 g_k 夹角。

$$m_k=-\theta_k^1 g_k+\beta_k^1 d_{k-1} \tag{9}$$

(8)式表明当 m_k 与 g_k 接近正交时，d_k 取负梯度方向。

下面引理表明，这种修正还带来了数值上的稳定性。

引理 1　$k \geqslant 1$ 且 $g_k \neq 0$ 时，$|\cos \eta_k| \neq 0$ 的充要条件是 $\Delta_k \neq 0$。

证明　$k \geqslant 1$ 时，由(6)、(9)式，直接计算得

$$m_k^{\mathrm{T}} g_k=\left(-\theta_k^1 g_k+\beta_k^1 d_{k-1}\right)^{\mathrm{T}} g_k=-c\|g_k\|^2 \Delta_k \tag{10}$$

从而命题成立。　　证毕

2 算法及收敛性

在这一节中，以(8)式为搜索方向给出算法。

算法 1　步 1，给定初值。选取初始点 $x_0 \in \mathbf{R}^n$，$\varepsilon \geqslant 0$，$0<\rho<1$，$0 \leqslant \theta<0.5$，$c>0.5$，$0<\theta'<1$，$k=0$，选初始方向 $d_0=-g_0$；

步 2，检验终止条件。若 $\|g_k\| \leqslant \varepsilon$，则 $x^*=x_k$，停止迭代；

步 3，由(8)式确定搜索方向 d_k；

步 4，用 Wolfe 线搜索准则求步长 $\alpha_k>0$；

步5,计算新点。令 $x_{k+1}=x_k+\alpha_k d_k$,$k=k+1$,转步2。

从(8)式知,若 $d_k=-g_k$,则 $d_k^{\mathrm{T}}g_k=-g_k^{\mathrm{T}}g_k$;若 $d_k=m_k/\Delta_k$,结合(10)式有,$d_k^{\mathrm{T}}g_k=-cg_k^{\mathrm{T}}g_k$。这表明算法1满足下降性。为了证明收敛性,先给出并证明下面一些引理。下面的引理及证明可参考文献[2]。

引理2 假设目标函数 f 的梯度 $\nabla f(x)$ Lipschitz 连续,算法1中步长 α_k 由 Wolfe 线搜索确定。若算法1产生的序列 $\{x_k\}$ 使得 $f(x_k)$ 有下界,则有

$$\sum\frac{(g_k^{\mathrm{T}}d_k)^2}{\|d_k\|^2}<\infty \tag{11}$$

即 Zoutendijk 条件成立。

下面引理表明,d_k 与 $-g_k$ 的夹角余弦有下界。

引理3 算法1中的搜索方向 d_k 满足:存在常数 $m>0$,使得对一切 $k\geqslant 0$,有 $\frac{(d_k^{\mathrm{T}}g_k)^2}{\|d_k\|^2}\geqslant m\|g_k\|^2$。

证明 若 $k=0$ 或 $|\cos\eta_k|<\theta'$,则 $d_k=-g_k$,$\frac{(d_k^{\mathrm{T}}g_k)^2}{\|d_k\|^2}=\|g_k\|^2$。否则 $d_k=\frac{m_k}{\Delta_k}$,由(10)式有

$$\|d_k\|^2=\left\|\frac{m_k}{\Delta_k}\right\|^2=\left(\frac{|m_k^{\mathrm{T}}g_k|}{|\cos\eta_k|\,\|g_k\|\,|\Delta_k|}\right)^2=\left(\frac{c}{|\cos\eta_k|}\|g_k\|\right)^2\leqslant\left(\frac{c}{\theta'}\|g_k\|\right)^2 \tag{12}$$

由(5)、(12)式,得 $\frac{(d_k^{\mathrm{T}}g_k)^2}{\|d_k\|^2}\geqslant\theta'^2\|g_k\|^2$。因此,结合 $0<\theta'<1$,只要令 $m=\theta'^2$,即有对一切 $k\geqslant 0$,$\frac{(d_k^{\mathrm{T}}g_k)^2}{\|d_k\|^2}\geqslant m\|g_k\|^2$。

证毕

定理1 假设目标函数 f 的梯度 $\nabla f(x)$ Lipschitz 连续,若算法1中步长 α_k 由 Wolfe 线搜索确定,那么算法1中要么存在某个 x_k 使得 $\nabla f(x_k)=0$,要么产生一个序列 $\{x_k\}$,使得

$$\liminf_{k\to\infty}\|\nabla f(x_k)\|=0 \text{ 或 } \liminf_{k\to\infty}f(x_k)=-\infty$$

证明 用反证法。假设定理不成立,则算法1产生一个序列 $\{x_k\}$ 使得 $f(x_k)$ 有下界且存在 $\delta>0$ 使得对一切 k 都有 $\|g_k\|\geqslant\delta$。

由于 $f(x_k)$ 有下界,又目标函数 f 的梯度 $\nabla f(x)$ Lipschitz 连续,根据引理2有(11)式成立。根据引理3,存在常数 $m>0$,使得对一切 $k\geqslant 0$,有 $\frac{(g_k^{\mathrm{T}}d_k)^2}{\|d_k\|^2}\geqslant m\|g_k\|^2\geqslant m\delta^2$,从而 $\sum\frac{(g_k^{\mathrm{T}}d_k)^2}{\|d_k\|^2}>\infty$。这与(11)式矛盾,得证。

证毕

3 数值实验

由于采取精确线搜索,取 $c=1$ 时,$\theta_k=1$,$\beta_k=\beta_k^{\mathrm{HS}}$,所以新方法可以看成是 HS 方法的推广。因此下面通过12个例子[13],将算法1和文献[9]中的 HS 型谱共轭梯度法(记为 SHS)进行数值比较实验。实验均在 PC 机上完成,PC 机的配置如下:Intel(R) Core(TM)i3-2350M CPU @ 2.30GHz,4GB 内存,Win7 系统。程序用 Matlab 编写,运行环境为 Matlab R2008a。线搜索统一使用文献[1]中的 Lines 程序,具体参数选取为 $\varepsilon=10^{-5}$,$\rho=0.01$,$\theta=0.01$,$c=0.99$,$\theta'=10^{-3}$。实验结果见表1,其中 CPU 表示计算时间,N_i 表示迭代次数,N_f 表示函数值计算次数,$\|g\|$ 表示计算完成时的梯度范数。

表1 算法1和 HS 型谱共轭梯度法[9]数值比较

测试函数	x_0	维数	CPU	N_i/N_f	$\|g\|$	算法
Hager	(1,1,…,1)	10	0.015 6	48/56	6.362 2e−006	算法1
			0.046 8	78/81	6.185 8e−006	SHS
Generalized White & Holst	(−1.2,1,…,−1.2,1)	10	0.078 0	548/1 571	4.679 2e−006	算法1
			0.093 6	638/1 838	8.689 4e−006	SHS
Extended Maratos	(1.1,0.1,…,1.1,0.1)	100	0.015 6	108/279	5.647 5e−007	算法1
			0.046 8	138/348	9.382 7e−007	SHS

续表 1

测试函数	x_0	维数	CPU	N_i/N_f	$\|g\|$	算法
FLETCHCR	(0,0,…,0)	100	0.187 2	1 138/3 401	9.791 2e−006	算法 1
			0.202 8	1 400/4 167	9.226 0e−006	SHS
Diagonal 7	(1,1,…,1)	1 000	0.015 6	8/14	5.947 1e−007	算法 1
			0.046 8	10/16	4.446 1e−008	SHS
		5 000	0.015 6	10/15	7.581 5e−007	算法 1
			0.046 8	10/16	3.555 6e−007	SHS
Extended Tridiagonal 1	(2,2,…,2)	1 000	0.046 8	202/348	3.049 5e−006	算法 1
			0.109 2	454/773	5.694 7e−006	SHS
		10 000	0.218 4	202/348	9.643 4e−006	算法 1
			1.466 4	1 502/2 718	1.335 7e−005	SHS
Diagonal 5	(1.1,1.1,…,1.1)	1 000	0.015 6	10/11	8.020 6e−008	算法 1
			0.046 8	38/39	4.158 7e−006	SHS
		10 000	0.031 2	6/8	6.774 8e−006	算法 1
			0.140 4	40/42	2.927 0e−006	SHS
Generalized PSC1	(3,0.1,…,3,0.1)	1 000	0.187 2	688/738	9.844 8e−006	算法 1
			0.327 6	1 102/1 406	9.627 6e−006	SHS
		10 000	1.700 4	799/979	1.189 0e−005	算法 1
			2.854 8	1 319/1 694	1.267 6e−005	SHS
Raydan 2	(1,1,…,1)	1 000	0.015 6	18/19	2.128 6e−008	算法 1
			0.031 2	66/67	9.783 8e−006	SHS
		10 000	0.015 6	14/15	1.960 4e−007	算法 1
			0.078 0	70/72	8.794 9e−007	SHS
Extended Powell	(3,−1,0,1,…,3,−1,0,1)	1 000	0.031 2	104/218	8.102 9e−006	算法 1
			0.062 4	228/466	2.782 1e−006	SHS
		10 000	0.249 6	218/434	3.284 4e−006	算法 1
			0.358 8	334/685	5.038 9e−006	SHS
Extended EP1	(1.5,1.5,…,1.5)	1 000	0.015 6	10/27	8.229 8e−006	算法 1
			0.031 2	11/62	8.959 2e−005	SHS
		10 000	0.046 8	8/22	7.798 1e−007	算法 1
			0.078 0	11/63	2.833 2e−004	SHS
Extended Hiebert	(0,0,…,0)	1 000	0.015 6	4/9	0	算法 1
			0.452 4	1 151/5 113	55.976 4	SHS
		10 000	0.031 2	4/9	0	算法 2.1
			3.369 6	1 502/6 866	5.338 8	SHS

从表 1 可以看出,对于所选测试函数,无论是迭代次数还是计算时间,算法 1 的数值结果都明显优于 SHS 算法。实际实验中发现,参数 θ、c 以及 θ' 的选取对实验结果影响较大,因此上述 3 个参数的(自适应)选取值得进一步研究。

参考文献：

[1] 倪勤.最优化方法及程序设计[M].北京：科学出版社，2009.
Ni Q. Optimization methods and procedures[M]. Beijing: Science Press, 2009.

[2] 戴彧虹，袁亚湘.非线性共轭梯度法[M].上海：上海科学出版社，2000.
Dai Y H, Yuan Y X. Nonlinear conjugate gradient method [M]. Shanghai: Shanghai Science Press, 2000.

[3] Birgin E G, Martinez J M. A spectral conjugate gradient method for unconstrained optimization[J]. Appl Math Optimiz, 2001, 43: 117-128.

[4] Zhang L, Zhou W, Li D. Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search[J]. Numer Math, 2006, 104(4): 561-572.

[5] Du S Q, Chen Y Y. Global convergence of a modified spectral FR conjugate gradient method[J]. Applied Mathematics and Computation, 2008, 202(2): 766-770.

[6] Lu A Q, Liu H M, Zheng X Y, et al. A variant spectral-type FR conjugate gradient method and its global convergence [J]. Applied Mathematics and Computation, 2011, 217(12): 5547-5552.

[7] 王开荣，曹伟，王银河. Armijo 型线搜索下的谱 CD 共轭梯度法[J].山东大学学报：理学版，2011，45(11)：104-108.
Wang K R, Cao W, Wang Y H. A spectral CD conjugate gradient method with Armijo-type line search[J]. Journal of Shandong University: Natural Science, 2011, 45(11): 104-108.

[8] 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.

[9] Du X L, Liu J K. Global convergence of a spectral HS conjugate gradient method[J]. Procedia Engineering, 2011, 15: 1487-1492.

[10] Wan Z, Yang Z L, Wang Y Y. New spectral PRP conjugate gradient method for unconstrained optimization[J]. Applied Mathematics Letters, 2011, 24(1): 16-22.

[11] 黄海，林穗华.一个 PERP 共轭型梯度法的收敛性[J].西南大学学报：自然科学版，2012，34(3)：22-29.
Huang H, Lin H H. The convergence property of a PRP-type conjugate gradient method[J]. Journal of Southwest Univerisity: Natural Science, 2012, 34(3): 22-29.

[12] Dai Y H, Liao L Z. New conjugacy conditions and related nonlinear conjugate gradient methods[J]. Appl Math Optim, 2001, 43(1): 87-101.

[13] Andrei N. An unconstrained optimization test functions collection[J]. Advanced Modeling and Optimization, 2008, 10 (1): 147-161.

Operations Research and Cybernetics

A Sufficient Descent Spectral Conjugate Gradient Method

CHEN Long-wei[1,2], XIA Fu-quan[3], JIA Chao-yong[3]

(1. College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing 210016;
2. Taizhou Higher Vocational School of Mechanical & Electrical Technology, Taizhou Jiangsu 225300;
3. Dept. of Mathematics and Physics, Bengbu College, Bengbu Anhui 233030, China)

Abstract: First, a class of sufficiently descent spectral conjugate gradient method is put forward, which satisfies both conjugacy condition and descent condition, while the standard conjugate gradient method only meets conjugacy condition. Then, the global convergence of the new method is proved with the reduction to absurdity under the Wolfe line search. Finally, iterative times and computing time are compared between the new algorithm and the existing SHS algorithm in twelve examples. The comparison results show that the new algorithm is superior to the SHS algorithm in these two aspects. The global convergence and the numerical superiority indicate that the new algorithm is an effective algorithm which is worth studying.

Key words: unconstrained optimization; spectral conjugate gradient method; sufficient descent condition; conjugate condition; global convergence

（责任编辑 黄 颖）