第 43 卷 第 6 期 河南师范大学学报(自然科学版) Vol.43 No.6
2015 年 11 月 Journal of Henan Normal University(Natural Science Edition) Nov. 2015

文章编号:1000-2367(2015)06-0025-05 DOI:10.16366/j.cnki.1000-2367.2015.06.004

一类 Armijo 搜索下新的共轭梯度法及其全局收敛性

董晓亮[1],杨喜美[2],黄元元[3]

(1.北方民族大学 数学与信息科学学院,银川 750021;2.河南师范大学 数学与信息科学学院,河南 新乡 453007;
3.河南科技大学 数学与统计学院,河南 洛阳 471023)

摘 要:为有效求解大规模无约束优化问题,提出了一类新的混合共轭梯度法.该方法在每步迭代中都不依赖于函数的凸性和搜索条件而自行产生充分下降方向.在适当的条件下,获证了在 Armijo 搜索下,即使求解非凸函数极小化的问题,算法也具有全局收敛性.同时,数值实验表明所提算法可以有效求解优化测试问题.

关键词:共轭梯度法;全局收敛性;充分下降条件;Armijo 搜索

中图分类号:O224 **文献标志码**:A

共轭梯度法是求解无约束优化问题$\min\limits_{x\in \mathbf{R}^n} f(x)$ 的一类重要方法,其迭代公式为:

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

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

其中,$g_k = \nabla f(x_k)$,d_k 是搜索方向,α_k 是步长,β_k 是参数.著名的方法有 FR 方法、PRP 方法、DY 方法和 HS 方法.

充分下降条件在下降算法的收敛性分析中的作用很重要,也即下式成立

$$-d_k^{\mathrm{T}} g_k \geq c_1 \| g_k \|^2 (c_1 > 0 \text{ 是某个常数}). \tag{3}$$

近年来,学者们构造出不依赖于任何线搜索可自行产生充分下降方向,如 Hager 和 Zhang 提出了著名的 HZ 共轭梯度法[1],其中 $\beta_k^{HZ} = \beta_k^{HS} - 2\dfrac{\| y_{k-1} \|^2}{(d_{k-1}^{\mathrm{T}} y_{k-1})^2} g_k^{\mathrm{T}} d_{k-1}$.它在 Wolfe 搜索下对非凸问题全局收敛,且总满足充分下降条件 $-d_k^{\mathrm{T}} g_k \geq \dfrac{7}{8} \| g_k \|^2$.

但 HZ 方法在 Arjimo 搜索下是否收敛目前是未知的.近来,Zhang 和 Li[2] 提出了一类 Armijo 搜索下 CHZ 方法,其中 $d_k^{CHZ} = \begin{cases} -g_k, & s_k^{\mathrm{T}} y_k < \varepsilon_1 \| gk \|^r \| \mid s_k \|^2, \\ -g_k + \beta_k^{\mathrm{HZ}} d_{k-1} & s_k^{\mathrm{T}} y_k \geq \varepsilon_1 \| g_k \|^r \| s_k \|^2. \end{cases}$

在此基础上,本文继续研究 Arjimo 搜索下 HZ 方法的收敛性.以下是本文的出发点.

Zhang[3] 等改进了标准的割线方程 $\nabla^2 f(x_k) s_{k-1} = y_{k-1}$,将该方程中 y_{k-1} 修改为 $w_{k-1} = y_{k-1} + \dfrac{\theta_{k-1} u}{s_{k-1}^{\mathrm{T}} u}$,$\theta_{k-1} = 3(2f_{k-1} - 2f_k + g_{k-1}^{\mathrm{T}} s_{k-1} + g_k^{\mathrm{T}} s_{k-1})$.同时,Zhang 还证明了若 f 充分光滑,当 $\| s_{k-1} \|$ 充分小时,则

$$\begin{cases} s_{k-1}(\nabla^2 f(x_k) s_{k-1} - y_{k-1}) = O(\| s_{k-1} \|^3), \\ s_{k-1}(\nabla^2 f(x_k) s_{k-1} - w_{k-1}) = O(\| s_{k-1} \|^4). \end{cases}$$

为了提高割线方程的逼近精度,Babaie-Kafaki[4] 进一步更改为

收稿日期:2014-09-03;**修回日期**:2015-07-26.

基金项目:国家自然科学基金(11361001);北方民族大学校级科研基金(2014XBZ09);北方民族大学基本科研项目(2015JBK419).

第 1 作者简介(通信作者):董晓亮(1981-),男,甘肃静宁人,北方民族大学讲师,研究方向为数学规划,E-mail:dongxl@nun.edu.cn.

$$z_{k-1}=\begin{cases}y_{k-1}, & \|s_{k-1}\|\geq 1\\ y_{k-1}+\dfrac{\max(\theta_{k-1},0)}{s_{k-1}^{\mathrm{T}}u}u, & \|s_{k-1}\|<1.\end{cases}\tag{4}$$

近年来,文献[5－6]研究了几类具自适应共轭梯度法,但Armijo搜索下它们的收敛性未知,且无法直接推广到 PRP 和 LS 等方法. 本文在文献[7－8]的基础上,提出如下混合算法:

$$d_k=\begin{cases}-g_k+\beta_k^{PRP}+d_{k-1}, g_k^{\mathrm{T}}d_{k-1}\leq 0,\\ -g_k+\beta_{k+1}^{MHZ}d_{k-1}, g_{k+1}^{\mathrm{T}}d_k>0,\end{cases}\tag{5}$$

其中 $\beta_k^{PRP+}=\dfrac{\max\{0,g_k^{\mathrm{T}}y_{k-1}\}}{\|g_{k-1}\|^2}$,以及 $\beta_k^{MHZ}=\dfrac{g_k^{\mathrm{T}}z_{k-1}}{d_{k-1}^{\mathrm{T}}z_{k-1}}-2\dfrac{\|z_{k-1}\|^2}{(d_{k-1}^{\mathrm{T}}z_{k-1}^2)^2}g_k^{\mathrm{T}}d_{k-1}$.

注 1　该方法可推广为 HZ-LS$^+$ 方法,可类似证明方法的全局收敛性,其中

$$d_k=\begin{cases}-g_k-\dfrac{\max\{0,g_k^{\mathrm{T}}y_{k-1}\}}{g_{k-1}^{\mathrm{T}}d_{k-1}}, g_k^{\mathrm{T}}d_{k-1}\leq 0,\\ -g_k+\beta_{k+1}^{MHZ}d_{k-1}, g_{k+1}^{\mathrm{T}}d_k>0.\end{cases}$$

全文内容安排如下:第 2 节叙述本文的算法,第 3 节证明算法的全局收敛性,第 4 节给出相关的数值实验,通过与 HZ 等方法的实验结果来对比、验证本文算法的有效性.

1　算法及其充分下降条件

基于第一节提出的搜索方向(5),本节给出新的共轭梯度法,约记为 *MHZ* 方法:

算法 1

步骤 1　选定初始步 x_1,参数 $\delta\in(0,1)$,$\rho(0,1)$ 及 $\varepsilon>0$,置 $k=1$;

步骤 2　计算 $g_k=\nabla f(x_k)$. 若 $\|g_k\|<\varepsilon$,则停止计算;否则由(5)式计算 d_k;

步骤 3　按照 Armijo 搜索计算步长 α_k:即求 $\alpha_k=\max\{\rho^j,j=0,1,2,\cdots\}$,

$$f(x_k+\alpha_kd_k)-f(x_k)\leq-\delta\alpha_k^2g_k^{\mathrm{T}}d_k;\tag{6}$$

步骤 4　计算 $x_{k+1}=x_k+\alpha_kd_k$,置 $k=k+1$,转步骤 2.

下面,给出一个引理,说明算法 1 中的搜索方向 d_k 可满足充分下降条件.

引理 1　若 d_k 由(5)式定义,且 $d_{k-1}^{\mathrm{T}}z_{k-1}\neq 0$,则 d_k 满足充分下降条件 $-d_k^{\mathrm{T}}g_k\geq\dfrac{7}{8}\|g_k\|^2$.

证明　在(2)式两边同时与 g_k 做内积得 $g_k^{\mathrm{T}}d_k=-\|g_k\|^2+\beta_kg_k^{\mathrm{T}}g_{k-1}$. 若 $g_k^{\mathrm{T}}d_{k-1}\leq 0$,注意此时 $\beta_k=\beta_k^{PRP+}\geq 0$,则有 $g_k^{\mathrm{T}}d_k\leq-\|g_k\|^2$. 若 $g_k^{\mathrm{T}}d_{k-1}>0$,则类似文献[1]中定理 1.1 可证.

2　算法的收敛性分析

本节证明算法 1 的全局收敛性. 这里总是假设存在 $\varepsilon>0$,使得 $\forall k\in\mathbf{N}$,有 $\|g_k\|>\varepsilon$ 成立. 否则,$f(x)$ 的一个稳定点已经找到. 同时,本文作如下必要的假设 H.

(H1) $f(x)$ 在水平集 $L_0=\{x\in\mathbf{R}^n\mid f(x)\leq f(x_1)$ 有界;(H2)$g(x)$ 在水平集 L_0 内连续可微且满足 Lipschitz 条件,即存在 $L>0$,使 $\|g(x)-g(y)\|\leq L\|x-y\|$,$x,y\in L_0$.

由假设(H1)和(H2)知,存在 $\exists\gamma>0,B>0$,满足 $\forall x\in L_0$,$\|g(x)\|\leq\gamma$,$\|x\|\leq B$.

引理 2[4]　若假设 H 成立,z_k 由(4)式定义,则 $\|z_k\|\leq 4L\|s_k\|$.

定理 1　若存在 $\varepsilon>0$,使得 $\forall\in\mathbf{N}$,有 $\|g_k\|>\varepsilon$ 成立,则存在 $K>0$,满足 $\|d_k\|\leq K$.

证明　首先对 β_k 进行估计. 若 $g_k^{\mathrm{T}}d_{k-1}\leq 0$,则有 $|\beta_k^{PRP}|=\dfrac{\|g_k\|\|y_{k-1}\|}{\|g_{k-1}\|}\leq\dfrac{\gamma L\alpha_{k-1}\|\|d_{k-1}\|}{\varepsilon^2}$. 若 $g_k^{\mathrm{T}}d_{k-1}>0$,则有 $-g_{k-1}^{\mathrm{T}}d_{k-1}\leq d_k^{\mathrm{T}}y_{k-1}$. 结合(4)和引理 1 得到

$$z_{k-1}^{\mathrm{T}}d_{k-1}>y_{k-1}^{\mathrm{T}}d_{k-1}>-g_{k-1}^{\mathrm{T}}d_{k-1}>\frac{7}{8}\|g_{k-1}\|^2>\frac{7}{8}\varepsilon^2.\tag{7}$$

由(7)和引理 2 可以估计 $|\beta_k^{MHZ}|$ 的上界:

$$|\beta_k^{MHZ}| \leq \left|\frac{g_k^{\mathrm{T}} z_{k-1}}{d_{k-1}^{\mathrm{T}} z_{k-1}} + 2 g_k^{\mathrm{T}} d_{k-1}\left(\frac{\| z_{k-1} \|}{d_{k-1}^{\mathrm{T}} z_{k-1}}\right)^2\right| \leq \frac{\| g_k \| \| z_{k-1} \|}{z_{k-1}^{\mathrm{T}} d_{k-1}} + 2\frac{\| z_{k-1} \|^2}{z_{k-1}^{\mathrm{T}} d_{k-1}} \leq \frac{32}{7}\frac{L\gamma \| s_{k-1} \|}{\varepsilon^2} + \frac{512}{7}\frac{L^2 B \| s_{k-1} \|}{\varepsilon 2}. \tag{8}$$

记 $M = \frac{32}{7}\frac{L\gamma}{\varepsilon^2} + \frac{512}{7}\frac{L^2 B}{\varepsilon^2}$，由(5)式推知 $\| d_k \| \leq \gamma + M\alpha_{k-1} \| d_{k-1} \|^2$. 其次，对(6)式两边关于 k 求和，由假设 H1 和 Armijo 搜索知 $\sum_{k\geq 1} \| \alpha_k d_k \|^2 < \infty$，则有 $\lim_{k\to\infty} \alpha_k \| d_k \| = 0$. 所以必 $\exists r(0,1)$ 及 $k_0 \in \mathbf{N}$，当 $k > k_0$ 时，满足 $M\alpha_{k-1} \| d_{k-1} \| < r$.

基于上述不等式，对 $\| d_k \| \leq \gamma + M\alpha_{k-1} \| d_{k-1} \|^2$ 进行递推，则有

$$\| d_k \| \leq \gamma + r \| d_{k-1} \| \leq \gamma \sum_{i=0}^{k-k_0-1} r^i + r^{k-k_0} \| d_{k_0} \| \leq \frac{\gamma}{1-r} + r^{k-k_0} \| d_{k_0} \|.$$

令 $K = \max\{ \| d_1 \|, \| d_2 \|, \cdots, \| d_{k_0} \|, \frac{\gamma}{1-r} + \| d_{k_0} \| \}$，可知 $\| d_k \| < K$.

类似于文献[3]中定理 3.2，可以证明本文算法的全局收敛性，在此略去证明过程.

定理 2 若假设 H 成立，d_k 由算法 1 确定，α_k 由 Armijo 搜索获得，则 $\liminf_{k\to\infty} \| g_k \| = 0$.

3 数值实验

本节，我们报告数值结果. 用文献[9]中的测试函数来检验本文的算法，所有的算法用 MATLAB 语言编写程序，结果如表 1 所示，其中各列的意义如下.

“HZ”表示 Wolfe 搜索下的 HZ 方法，“CHZ”表示采用 Armijo 搜索的 CHZ 方法，“$MHZ1$”和“$MHZ2$”为 Armijo 搜索下的本文两类方法. $MHZ1$ 方法和 $MHZ2$ 方法分别取 $u = y_{k-1}$ 和 $u = s_{k-1}$. “P”代表算例名称，“Dim”代表算例规模，“NI”表示迭代次数，“NF”表示目标函数的计算次数，“NG”表示梯度的计算次数，“TIME”表示单位为 s 的 CPU 运行时间.

下面就一些参数进行说明，HZ 算法的步长由强 Wolfe 搜索获得，即 α_k 满足

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \lambda \alpha_k g_k^{\mathrm{T}} d_k, \ | g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k | \leq \sigma | g_k^{\mathrm{T}} d_k |.$$

在实际求解中，取 $\lambda = 0.01$，$\sigma = 0.01$.

对于后三者算法，本文采用统一的 Armijo 搜索求解步长，在(6)式中，$\delta = 10^{-4}$，$\rho = 0.5$. 文献[10]给出了 Armijo 搜索中一类 自调比的初始步长 $\alpha_{k0} = \frac{\varepsilon g_k^{\mathrm{T}} d_k}{(g(x_k) - g x_k + \varepsilon d_k())^{\mathrm{T}}}$ 其中 $\varepsilon = 10^{-8}$. 在实际求解中，若初始步长出现 $\alpha_{k0} < 5\times 10^{-6}$，调整为 $\alpha_{k0} = 1$. CHZ 算法[2]：$r = 2$，以及当 $\| g(x_k) \| < 100$ 时，$\varepsilon_1 = 0.1$；否则，$\varepsilon_1 = 10^{-6}$.

表 1 数值实验结果对比

P	Dim	HZ	$MHZ1$	$MHZ2$	CHZ
		NI/NF/NG/TIME	NI/NF/TIME	NI/NF/TIME	NI/NF/TIME
ROSE	2	32/133/105/0.032	39/117/0.033	32/103/0.015	32/90/0.013
FROTH	2	14/67/54/0.012	11/23/0.022	11/23/0.0047	16/33/0.0065
BEALE	2	11/41/30/0.097	17/35/0.0082	17/35/0.0073	20/41/0.008
JENSAM	2	9/51/36/0.029	9/20/0.0044	11/24/0.0053	11/25/0.0051
HELIX	3	32/100/77/0.071	49/102/0.022	40/93/0.019	105/217/0.043
BARD	3	39/117/92/0.033	50/101/0.03	29/59/0.018	42/86/0.024
GAUSS	3	4/9/5/0.018	3/7/0.0021	3/7/0.002	3/7/0.0021
MEYER	3	1000/5532/5029/1	3/109/0.0083	3/109/0.0093	5/113/0.009
GULF	3	1/2/2/0.045	1/3/0.0012	1/3/0.0011	1/3/0.001

(续表)

P	Dim	HZ	MHZ1	MHZ2	CHZ
		NI/NF/NG/TIME	NI/NF/TIME	NI/NF/TIME	NI/NF/TIME
BOX	3	58/183/155/0.048	50/103/0.027	43/89/0.025	75/153/0.038
SING	4	83/267/226/0.046	418/837/0.21	169/339/0.07	667/1335/0.33
WOODP	4	92/254/205/0.099	156/335/0.081	133/295/0.065	201/424/0.12
KOWOSB	4	96/265/237/0.066	100/202/0.064	69/140/0.039	110/225/0.11
BIGGS	6	220/683/597/0.16	982/1993/0.62	142/297/0.11	60/127/0.068
OSB2	11	210/543/481/0.4	100/214/0.15	227/457/0.36	266/536/0.41
	100	28/120/99/0.078	40/119/0.056	37/122/0.037	61/143/0.047
ROSEX	500	31/140/114/0.61	37/113/0.51	37/113/0.48	33/100/0.41
	1000	69/226/192/4	50/181/3.4	127/255/4	534/1069/17
	100	44/137/112/0.045	628/1257/0.49	138/277/0.11	92/191/0.089
SINGX	500	37/221/171/3	47/159/1.8	37/107/1.2	47/141/1.6
	1000	38/199/151/6.9	48/157/5	33/102/3.3	42/110/3.8
PEN1	20	34/209/156/0.12	45/141/0.065	35/101/0.04	39/82/0.045
	50	30/217/157/0.15	119/535/0.34	93/395/0.24	43/179/0.12
PEN2	20	448/1629/1418/0.38	79/188/0.055	127/290/0.093	223/503/0.15
	40	632/1973/1707/0.64	118/249/0.12	126/268/0.11	206/439/0.18
VARDIM	100	12/158/103/0.072	4/161/0.022	4/161/0.022	4/161/0.033
	500	14/209/138/0.86	3/108/0.27	3/108/0.28	4/161/0.42
	100	50/135/116/0.37	51/110/0.29	50/108/0.28	50/108/0.28
TRIG	300	43/113/97/12	51/109/12	47/100/11	50/108/11
	500	49/126/107/98	49/102/84	52/108/86	50/106/83
	100	344/689/391/0.43	344/689/0.42	344/689/0.41	344/689/0.42
BV	500	14/23/22/0.12	14/29/0.17	14/29/0.17	14/29/0.17
	1000	0/1/1/0.017	0/1/0.033	0/1/0.023	0/1/0.023
	100	6/13/8/0.084	6/13/0.089	6/13/0.09	6/13/0.1
IE	500	7/16/10/2	7/15/2.4	7/15/2.4	6/13/2
	1000	7/16/10/8	7/15/9.2	7/15/9.2	7/15/9.2
	100	28/61/35/0.035	27/55/0.023	27/55/0.026	26/53/0.033
TRID	500	31/70/41/0.26	29/59/0.28	29/59/0.28	27/55/0.25
	1000	32/74/41/0.97	31/63/1.1	31/63/1.1	28/57/0.95
	100	18/154/47/0.18	17/35/0.036	17/35/0.048	19/39/0.04
BAND	500	21/270/46/4.5	16/33/0.49	16/33/0.5	19/39/0.57
	1000	21/245/57/15	17/35/2	17/35/2	19/39/2.2
LIN	100	1/8/8/0.034	1/3/0.006 9	1/3/0.006 6	1/3/0.007 9
	500	1/9/9/0.3	1/3/0.11	1/3/0.11	1/3/0.1

另外,统计了上述 4 类算法的迭代总体信息,如表 2 所示,其中 HZ 方法只统计 NF.

表 2 4 类算法的总体比较

Algorithm	HZ	MHZ1	MHZ2	CHZ
Total NI	2991	3871	2390	3604
Total NF	10 176	8829	5627	7951
Total TIME/s	160.311	125.9008	124.297	135.4166

综上所述,本文算法在一定程度上可与 HZ 方法相媲美,说明它是有效的.

参 考 文 献

[1] Hager W W, Zhang H. A new conjugate gradient method with guaranteed descent and an efficient line search[J]. SIAM J Optim, 2005, 16(1): 170-192.

[2] 张 丽,周伟军. Armijo 搜索下 Hager-Zhang 共轭梯度法的全局收敛性[J]. 数学物理学报, 2008, 28A(5): 840-845.

[3] Zhang J, Xu C. Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations[J]. J Comp Appl Math, 2001, 137(2): 269-278.

[4] Babaie-Kafaki S, Ghanbari R. Two new conjugate gradient methods based on modified secant equations[J]. J Comp Appl Math, 2010, 234

(1):1374-1386.

[5] Dong X, Liu H, He Y. A self-adjusting conjugate gradient method with sufficient descent condition and conjugacy condition[J]. J Optim Theory Appl, 2015, 165(1):225-241

[6] Dong X, Liu H, He Y, et al. A modified Hestenes-Stiefel conjugate gradient method with sufficient descent condition and conjugac condition[J]. J Comp Appl Math, 2015, 281:239-249.

[7] Dong X, Liu H, He Y. New version of the three-term conjugate gradient method based on spectral scaling conjugacy condition that generates descent search direction[J]. Appl Math Comput, 2015, 269:606-617.

[8] 董晓亮，高岳林，何郁波. 一类混合的 FR-PC 共轭梯度法及其全局收敛性[J]. 河南师范大学学报(自然科学版), 2010, 35(2):42-44.

[9] MorèJ J, Garbow B S, Hillstrome K E. Testing unconstrained optimization software[J]. ACM Trans Math Software, 1981, 7(1):17-41.

[10] Zhang L, Zhou W, Li D. Adescent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence[J]. IMA J Numer Anal, 2006, 26(4):629-640.

Global Convergenceof a New Conjugate Gradient Method with Armijo Search

DONG Xiaoliang[1], YANG Ximei[2], HUANG Yuanyuan[3]

(1. School of Mathematics and Information, Beifang University of Nationalities, Yinchuan 710021, China;

2. College of Mathematics and Information Science, Henan Normal University, Xinxiang 453007, China;

3. School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang 471023, China)

Abstract: A new kind of hybrid conjugate gradient method for solving large scale unconstrained optimization problems is proposed. The modified method provides automatically a sufficient descent direction for the objective function at each iteration, a property depends neither on the line search used, nor on the convexity of the function. Under appropriate conditions, the proposed method with the Armijo line search converges globally even if the objective function is nonconvex. Numerical results show that the new method is efficient and can be used to deal with some test problems.

Keywords: global convergence; sufficient descent condition; Armijo line search