广西科学 Guangxi Sciences 2007,14(4):378～380,385

一种新的杂交共轭梯度算法*

A New Hybrid Conjugate Gradient Method

林穗华[1,2],黄　海[1]

LIN Sui-hua[1,2],HUANG Hai[1]

(1. 南宁师范高等专科学校数学与计算机科学系,广西龙州　532400;2. 华东师范大学统计系,上海　200062)

(1. Department of Mathematics and Computer Science, Nanning Teachers College, Longzhou, Guangxi, 532400, China; 2. Department of Statistics East China Normal University, Shanghai, 200062, China)

摘要:结合 HS、DY 和 WYL 方法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:$\beta_k=\frac{\max\{0,\min\{\|g_k\|^2,g_k^Ty_{k-1},g_k^T\hat{y}_{k-1}\}\}}{d_{k-1}^Ty_{k-1}}$,并给出新的杂交共轭梯度算法;证明新算法在弱 Wolf-Powell 线搜索条件下具有全局收敛性,并用数值试验表明新算法具有较好的数值结果.

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:O224　　**文献标识码**:A　　**文章编号**:1005-9164(2007)04-0378-03

Abstract: Combined with HS, DY and WYL methods, a new hybrid conjugate gradient method for unconstrained optimization is proposed. Its updated formula is $\beta_k=\frac{\max\{0,\min\{\|g_k\|^2,g_k^Ty_{k-1},g_k^T\hat{y}_{k-1}\}\}}{d_{k-1}^Ty_{k-1}}$. We prove that, the corresponding method can ensure the global convergence under weak Wolf-Powell line search. Preliminary numerical results show that the proposed method is very efficient.

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

非线性共轭梯度法是求解大规模无约束优化问题的最重要的有效方法之一.考虑无约束优化问题

$$\min\{f(x)\mid x\in R^n\},\tag{0.1}$$

其中 $f:R^n\to R$ 为一阶可微非线性函数,其梯度 $\nabla f(x)$ 记为 $g:R^n\to R^n$. 共轭梯度法的迭代公式为

$$x_{k+1}=x_k+t_kd_k,\tag{0.2}$$

其中 t_k 为步长因子,d_k 为搜索方向.

常用的弱 Wolf-Powell(WWP) 线搜索规则为,寻找一个 $t_k>0$ 满足以下两式:

$$f(x_k+t_kd_k)-f(x_k)\leqslant\delta t_kg_k^Td_k,\tag{0.3}$$

$$g(x_k+t_kd_k)^Td(x_k)\geqslant\sigma g_k^Td_k,\tag{0.4}$$

其中参数 $0<\delta<1,\sigma\in(\delta,1)$.

在(0.2)式中搜索方向 d_k 定义为

$$d_k=\begin{cases}-g_k,k=1,\\-g_k+\beta_kd_{k-1},k\geqslant 2,\end{cases}\tag{0.5}$$

其中 β_k 为标量. 文献[1～8]给出选取 β_k 的几种公式如下:

$$\beta_k^{HS}=\frac{g_k^Ty_{k-1}}{d_{k-1}^Ty_{k-1}},\tag{0.6}$$

$$\beta_k^{FR}=\frac{g_k^Tg_k}{g_{k-1}^Tg_{k-1}},\tag{0.7}$$

$$\beta_k^{PRP}=\frac{g_k^Ty_{k-1}}{g_{k-1}^Tg_{k-1}},\tag{0.8}$$

$$\beta_k^{CD}=-\frac{g_k^Tg_k}{g_{k-1}^Td_{k-1}},\tag{0.9}$$

$$\beta_k^{LS}=-\frac{g_k^Ty_{k-1}}{g_{k-1}^Td_{k-1}},\tag{0.10}$$

$$\beta_k^{DY}=\frac{g_k^Tg_k}{d_{k-1}^Ty_{k-1}},\tag{0.11}$$

$$\beta_k^{WYL}=\frac{g_k^T\hat{y}_{k-1}}{g_{k-1}^Tg_{k-1}},\tag{0.12}$$

其中 $y_{k-1}=g_k-g_{k-1}$, $\hat{y}_{k-1}=g_k-\frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}$.

共轭梯度法收敛性分析中常用的下降性条件为

收稿日期:2006-12-12

修回日期:2007-03-26

作者简介:林穗华(1973-),女,讲师,主要从事概率统计与优化算法的教学与研究。

* 南宁师范高等专科学校科研项目(2007012)资助。

$$g_k^T d_k < 0, \forall\, k \geqslant 1. \tag{0.13}$$

多年来 β_k 的这几种公式在某些线搜索下的特性已被广泛研究，PRP 方法是数值结果最好的方法之一，但其收敛性差. WYL 方法在强 Wolf-Powell 条件下具有全局收敛性[9]，且有很好的数值结果. 为了得到收敛条件低，数值结果好的算法，就出现了将这些公式结合起来的杂交共轭梯度法. 如 1990 年 Touati-Ahmed 和 Story[10] 结合 β_k^{FR} 和 β_k^{PRP} 首先引入杂交方法

$$\beta_k = \max\{0, \min\{\beta_k^{FR}, \beta_k^{PRP}\}\}. \tag{0.14}$$

1992 年 Gilbert-Nocedal[11] 进一步研究杂交方法

$$\beta_k = \max\{-\beta_k^{FR}, \min\{\beta_k^{FR}, \beta_k^{PRP}\}\}. \tag{0.15}$$

这两种方法比 FR 方法好，但仍比 PRP 方法差；2001 年 Dai-Yuan[12] 结合 β_k^{HS} 和 β_k^{DY} 提出杂交方法

$$\beta_k = \max\{-c\beta_k^{DY}, \min\{\beta_k^{HS}, \beta_k^{DY}\}\}, \tag{0.16}$$

$$\beta_k = \max\{0, \min\{\beta_k^{HS}, \beta_k^{DY}\}\}, \tag{0.17}$$

其中 $c = \dfrac{1-\sigma}{1+\sigma} > 0$，它们在 WWP 条件下全局收敛，且数值结果均优于 PRP 方法，特别是后者对解决较为困难的问题要比 PRP 方法好；2005 年 Wei-Zhao-Chen[13] 结合 β_k^{HS}、β_k^{FR}、β_k^{PRP}、β_k^{CD}、β_k^{LS}、β_k^{DY} 提出杂交方法

$$\beta_k = \frac{\min\{\|g_k\|^2, g_k^T y_{k-1}\}}{\max\{-g_{k-1}^T d_{k-1}, d_{k-1}^T y_{k-1}, \|g_{k-1}\|^2\}},$$

$$\beta_k^* = \max\{0, \beta_k\}, \tag{0.18}$$

其收敛性和数值结果均好于 PRP 方法. 本文基于文献[11,12]的思想，结合 HS、DY 和 WYL 方法提出一种新的杂交共轭梯度法 β_k 公式.

1 新的共轭梯度算法

新杂交共轭梯度法的

$$\beta_k = \frac{\max\{0, \min\{\|g_k\|^2, g_k^T y_{k-1}, g_k^T \hat{y}_{k-1}\}\}}{d_{k-1}^T y_{k-1}}. \tag{1.1}$$

新的共轭梯度算法(算法 1)的步骤如下.

步骤 1：给定初值 $x_1 \in R^n$，$\varepsilon \geqslant 0$，计算 g_1，$d_1 := -g_1$，$k := 1$. 若 $\|g_1\| \leqslant \varepsilon$，停止.

步骤 2：计算步长因子 t_k，使满足 WWP 线搜索条件(0.3)式和(0.4)式.

步骤 3：迭代计算 $x_{k+1} := x_k + t_k d_k$，$g_{k+1} := g(x_{k+1})$. 若 $\|g_{k+1}\| \leqslant \varepsilon$，停止.

步骤 4：由(1.1)式计算 β_k，由(0.5)式计算 d_{k+1}.

步骤 5：$k := k+1$，转步骤 2.

2 算法 1 的收敛性

首先给出收敛性分析中要用到的两个假设.

假设 1 $f(x)$ 在水平集 $\Omega = \{x \in R^n \mid f(x) \leqslant f(x_1)\}$ 有下界，其中 x_1 为初始点.

假设 2 在水平集 Ω 上梯度 g 满足 Lipschitz 条件，即存在常数 $L > 0$，使 $\forall\, x, y \in \Omega$，有

$$\|g(y) - g(x)\| \leqslant L\|y - x\|. \tag{2.1}$$

定理 2.1 考虑算法 1 生成的序列 $\{g_k\}$ 和 $\{d_k\}$，若 $\forall\, k \geqslant 1$ 有 $g_k \neq 0$，则对 $\forall\, k \geqslant 1$，下降性条件(0.13)式成立.

证明 由 $d_1 = -g_1$，可得 $g_1^T d_1 = -\|g_1\|^2 < 0$. 以下假设 $g_{k-1}^T d_{k-1} < 0$.

由 WWP 条件(0.4)式可得

$$d_{k-1}^T y_{k-1} \geqslant (\sigma - 1) g_{k-1}^T d_{k-1} > 0. \tag{2.2}$$

由(1.1)式和(2.2)式可知，在 WWP 条件下有

$$\beta_k^{DY} > 0, 0 \leqslant \beta_k \leqslant \beta_k^{DY}. \tag{2.3}$$

可设 $\beta_k = r_k \beta_k^{DY}$，$r_k \in [0,1]$， (2.4)

若 $r_k = 0$，则由(0.5)式和(2.4)式，可知 $d_k = -g_k$，从而有 $g_k^T d_k = -\|g_k\|^2 < 0$.

以下设 $0 < r_k \leqslant 1$，(0.5)式两边与 g_k 作内积可得

$$g_k^T d_k = -\|g_k\|^2 + r_k \beta_k^{DY} g_k^T d_{k-1} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}(-d_{k-1}^T y_{k-1} + r_k g_k^T d_{k-1}) = \beta_k^{DY}[g_{k-1}^T d_{k-1} + (r_k - 1) g_k^T d_{k-1}]. \tag{2.5}$$

又由 $r_k - 1 \leqslant 0$，及(0.4)式 $g_k^T d_{k-1} \geqslant \sigma g_{k-1}^T d_{k-1}$，可得

$$(r_k - 1) g_k^T d_{k-1} \leqslant (r_k - 1)\sigma g_{k-1}^T d_{k-1}. \tag{2.6}$$

由 $0 < r_k \leqslant 1$，$0 < \sigma < 1$，可知 $(r_k - 1)\sigma \geqslant r_k - 1$，从而

$$1 + (r_k - 1)\sigma \geqslant 1 + (r_k - 1) = r_k > 0. \tag{2.7}$$

由(2.6)式和(2.7)式，及 $g_{k-1}^T d_{k-1} < 0$ 可得

$$g_{k-1}^T d_{k-1} + (r_k - 1) g_k^T d_{k-1} \leqslant [1 + (r_k - 1)\sigma] g_{k-1}^T d_{k-1} \leqslant r_k g_{k-1}^T d_{k-1} < 0. \tag{2.8}$$

由(2.3)式、(2.5)式和(2.8)式可得

$$g_k^T d_k \leqslant \beta_k^{DY} r_k g_{k-1}^T d_{k-1} < 0.$$

所以，由数学归纳法可知 $\forall\, k \geqslant 1$ 有 $g_k^T d_k < 0$，定理 2.1 得证.

引理 2.1 若假设 1，假设 2 成立，考虑算法 1 生成的序列 $\{g_k\}$ 和 $\{d_k\}$，则 Zoutendijk 条件成立，即

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty. \tag{2.9}$$

证明 由定理 2.1，有 $g_k^T d_k < 0$.

由(0.4)式及(1.2)式，可得

$$-(1-\sigma) g_k^T d_k \leqslant (g_{k+1} - g_k)^T d_k \leqslant L t_k \|d_k\|^2. \tag{2.10}$$

由(2.9)式知 $t_k \geqslant \frac{\sigma-1}{L}\frac{g_k^T d_k}{\| d_k \|^2}$. (2.11)

由(0.3)式及(2.11)式,可得

$$f_k - f_{k+1} \geqslant -\delta t_k g_k^T d_k \geqslant \delta \frac{1-\sigma}{L}\frac{(g_k^T d_k)^2}{\| d_k \|^2}. \quad (2.12)$$

又由假设1知,$\{f_k\}$ 单调递减且有界,所以 $\lim\limits_{k\to\infty} f_{k+1} < +\infty$,对(2.12)式左、右两边分别求和,可得

$$+\infty > f_1 - \lim_{k\to\infty} f_{k+1} = \sum_{k\geqslant 1}(f_k - f_{k+1}) \geqslant \frac{\delta(1-\sigma)}{L}\sum_{k=1}^{\infty}\frac{(g_k^T d_k)^2}{\| d_k \|^2}.$$

所以(2.8)式成立,引理2.1得证.

定理2.2 若假设1,假设2成立,考虑算法1,则算法或者终止于稳定点或者

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

证明 若定理不成立,则存在 $q>0$,使 $\forall\, k\geqslant 1$ 有

$$\| g_k \| \geqslant q. \quad (2.13)$$

因 $\beta_k^{DY} = \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$,所以

$$\beta_k = r_k \beta_k^{DY} = r_k \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}. \quad (2.14)$$

又由 $d_k + g_k = \beta_k d_{k-1}$,可得

$$\| d_k \|^2 = \beta_k^2 \| d_{k-1} \|^2 - 2 g_k^T d_k - \| g_k \|^2. \quad (2.15)$$

(2.15)式两边除以 $(g_k^T d_k)^2$,利用(2.14)式及 $r_k \in [0,1]$,可得

$$\frac{\| d_k \|^2}{(g_k^T d_k)^2} = r_k^2 \frac{\| d_{k-1} \|^2}{(g_{k-1}^T d_{k-1})^2} - \frac{2}{g_k^T d_k} - \frac{\| g_k \|^2}{(g_k^T d_k)^2} = r_k^2 \frac{\| d_{k-1} \|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\| g_k \|^2} - \left(\frac{1}{\| g_k \|} + \frac{\| g_k \|}{g_k^T d_k}\right)^2 \leqslant r_k^2 \frac{\| d_{k-1} \|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\| g_k \|^2} \leqslant \frac{\| d_{k-1} \|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\| g_k \|^2}. \quad (2.16)$$

利用(2.16)式进行递推,并注意到 $\| d_1 \|^2 = -g_1^T d_1 = \| g_1 \|^2$,及由(2.13)式,可得

$$\frac{\| d_k \|^2}{(g_k^T d_k)^2} \leqslant \sum_{i=1}^{k}\frac{1}{\| g_k \|^2} \leqslant \frac{1}{q^2}k. \quad (2.17)$$

由(2.17)式可得

$$\frac{(g_k^T d_k)^2}{\| d_k \|^2} \geqslant \frac{q^2}{k}. \quad (2.18)$$

由(2.18)式可知

$$\sum_{k\geqslant 1}\frac{(g_k^T d_k)^2}{\| d_k \|^2} = +\infty. \quad (2.19)$$

(2.19)式与引理2.1矛盾,所以定理2.2得证.

与文献[13]中的分析类似,(1.1)式对应的杂交共轭梯度法能够避免小步长的倾向.

3 数值试验

为考察算法1的数值表现,分别对由(0.8)式、(0.17)式、(1.1)式计算 β_k,(0.2)式、(0.5)式迭代格式对应的共轭梯度法,用Matlab程序进行数值试验,试验不采用重开始技术,试验参数为 $\delta=0.01,\sigma=0.1,\varepsilon=10^{-5}$,结果见表1.

表1 PRPSWP、DYHybrid和NewHybrid的试验结果

Table 1 Test results for the PRPSWP/DYHybrid/NewHybrid

Proplem	Dim	NI/NF/NG PRPSWP	DYHybrid	NewHybrid
Rose	2	—	53/384/80	44/426/81
Froth	2	11/76/22	62/103/86	62/103/86
Badscp	2	—	—	—
Badscb	2	—	—	—
Beale	2	13/79/24	34/104/46	34/104/46
Jensam	2	10/169/19	23/93/37	23/93/37
Helix	3	35/231/66	25/46/38	47/184/79
Bard	3	30/64/54	44/79/64	41/71/59
Gauss	3	4/57/6	3/6/5	3/6/5
Meyer	3	—	—	—
Gulf	3	1/2/2	1/2/2	1/2/2
Box	3	—	—	—
Sing	4	105/367/177	55/103/84	118/320/179
Wood	4	108/544/206	100/354/166	107/354/168
Kowosb	4	112/376/185	35/168/58	35/121/59
Bd	4	—	27/109/48	36/120/59
Osb1	5	—	—	—
Biggs	6	127/512/214	—	129/360/200
Osb2	11	—	—	—
Watson	20	2765/6906/4329	868/1788/1282	637/1355/947
Rosex	8	26/472/76	36/263/85	46/330/105
	50	31/633/88	43/231/74	30/123/58
	100	—	47/402/88	53/175/97
Singx	8	206/999/355	80/256/127	124/333/184
Pen1	2	19/308/58	24/161/52	9/37/27
Pen2	4	13/42/30	15/56/32	18/107/35
	50	2214/6400/3583	141/557/251	102/422/188
Vardim	2	3/8/7	3/55/6	3/55/6
	50	9/32/25	11/43/36	11/43/36
Trig	3	14/268/26	18/83/28	14/30/23
	50	37/402/60	45/73/62	37/114/54
	100	46/331/88	51/233/75	49/83/73
Bv	3	14/27/24	15/24/20	15/24/20
	10	92/415/154	110/324/151	160/396/215
Ie	3	5/11/10	6/10/9	6/10/9
	50	5/10/9	6/9/8	6/9/8
	100	5/10/10	10/11/11	10/11/11
	200	5/10/10	6/7/7	6/7/7
	500	7/14/14	7/10/10	7/10/10
Trid	3	16/29/25	22/29/27	22/29/27
	50	26/139/41	48/50/50	48/50/50
	100	28/380/35	34/89/38	34/89/38
	200	30/191/42	46/102/51	46/102/51
Band	3	9/62/13	13/15/15	13/15/15
	50	18/131/31	20/174/29	20/174/29
	100	18/187/33	23/227/33	23/227/33
	200	19/184/34	20/226/31	20/226/31
Lin	2	1/3/3	1/3/3	1/3/3
	50	1/3/3	1/3/3	1/3/3
	500	1/3/3	1/3/3	1/3/3
	1000	1/3/3	1/3/3	1/3/3
Lin1	2	1/51/2	1/51/2	1/51/2
	10	1/3/3	1/3/3	1/3/3
Lin0	4	1/3/3	1/3/3	1/3/3

(下转第385页 Continue on page 385)

converges to 0, hence $\{x_n\}_{n=-2k-1}^{\infty}$ converges to $\frac{\alpha}{A-q}$.

The proof of Theorem 1.4 is completed.

References:

[1] WANG QI, ZENG F, ZHANG G, et al. On the global character of the difference Equation $x_{n+1}=(\alpha+B_1x_{n-1}+B_3x_{n-3}+\cdots+B_{2k+1}x_{n-2k-1})/(A+B_0x_n+B_2x_{n-2}+\cdots+B_{2k}x_{n-2k})$ [J]. Journal of Difference Equations and Applications, 2006, 12(5): 399-417.

[2] KULENOVI C M R S, LADAS G. Dynamics of second order rational difference equations with open problems and conjectures[M]. Chapman & Hall/CRC Press, 2001.

[3] LI W, SUN H. Dynamics of a rational difference Equation[J]. Appl Math Comput, 2005, 163: 577-591.

[4] CHATTERJEE E, GROVE E A, KOSTROV Y, et al. On the trichotomy character of $x_{n+1}=(\alpha+\gamma x_{n-1})/(A+Bx_n+x_{n-2})$ [J]. Journal of Difference Equations and Applications, 2003, 9(12): 1113-1128.

[5] GROVE E A, LADAS G, PREDESCU, et al. On the global character of the difference Equation $x_{n+1}=(\alpha+\gamma x_{n-(2k+1)}+\delta x_{n-2l})/(A+x_{n-2l})$ [J]. Journal of Difference Equations and Applications, 2003, 9(2): 171-199.

(责任编辑:邓大玉　蒋汉明)

(上接第 380 页 Continue from page 380)

表 1 中的相关符号意义为:Problem 是测试问题的名称;Dim 是目标函数的维数;NI 是算法迭代的次数;NF 是函数值计算的次数;NG 是函数梯度计算的次数;PRPSWP 是(0.8)式＋SWP;DYHybrid 是(0.17)式＋WWP;NewHybrid 是(1.1)式＋WWP.

表 1 的数据结果显示,对测试问题集的 54 个目标函数,PRPSWP、DYHybrid 和 NewHybrid 方法求解失败的个数分别为 9、7、6 个. NewHybrid 方法的 NI/NF/NG 数据优于 DYHybrid 和 PRPSWP 方法.

致谢:

本文在完成过程中得到广西大学韦增欣教授的悉心指导,他对本文提出宝贵的修改意见,我们谨此深表感谢!

参考文献:

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

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

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

[4] POLAK E, RIBIRE G. Note sur la convergence de directions conjugees [J]. Rev Francaise Informat Recherche Operatinelle 3e Annee, 1969, 16: 35-43.

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

[6] LIU Y, STORY C. Efficient generalized conjugate gradient algorithms Part 1: Theory[J]. JOTA, 1991, 69: 129-137.

[7] DAI Y, YUAN Y. A nonlinear conjugate gradient method with the strong global convergence properties [J]. SIAM Journal on Optimization, 1999, 10: 177-182.

[8] WEI Z X, YAO S W L, LIU L Y. The convergence properties of some new conjugate gradient methods[J]. Applied Mathematics and Computation, 2006, 183: 1341-1350.

[9] HUANG H, WEI Z X, YAO S W. The proof of the sufficient descent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search[J]. Applied Mathematics and Computation, 2007, 189: 1241-1245.

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

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

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

[13] WEI Z X, ZHAO Y, CHEN C L. A new hybrid conjugate gradient method for unconstrained optimization [J]. Journal of Qufu Normal University, 2005, 31(4): 35-39.

(责任编辑:尹　闯　邓大玉)