文章编号：1000-5471(2007)06-0043-05

一个双参数的共轭梯度法簇[①]

林穗华[1,2]，黄　海[2]

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

摘要： 提出了含有双参数的新共轭梯度法 β_k 公式，证明了该方法在适当选取参数 σ 的 SWP 线搜索下具有充分下降性和全局收敛性，数值试验结果良好.

关　键　词： 无约束最优化；共轭梯度法；SWP 线搜索；全局收敛

中图分类号： O224　　　　**文献标识码：** A

共轭梯度法适合求解大规模无约束优化问题 $\min\{f(x) \mid x \in R^n\}$，目标函数 $f: R^n \to R$ 为连续可微的非线性函数，$f(x)$ 的梯度记为 $g(x)$，f_k 表示 $f(x_k)$，g_k 表示 $g(x_k)$. 共轭梯度法具有如下形式：

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

其中：t_k 为步长因子，由线性搜索产生，d_k 为搜索方向，计算公式为：

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

其中 β_k 为标量参数. 文献[1-5]提出如下 β_k 公式：

$$\beta_k^{\mathrm{PRP}} = \frac{g_k^T y_{k-1}}{g_{k-1}^T g_{k-1}} \tag{3}$$

$$\beta_k^{\mathrm{FR}} = \frac{\| g_k \|^2}{\| g_{k-1} \|^2}$$

$$\beta_k^{\mathrm{WYL}} = \frac{g_k^T \hat{y}_{k-1}}{g_{k-1}^T g_{k-1}} \tag{4}$$

$$\beta_k^{\mathrm{MHS}} = \frac{g_k^T \hat{y}_{k-1}}{d_{k-1}^T y_{k-1}} \tag{5}$$

$$\beta_k^{\mathrm{MLS}} = \frac{g_k^T \hat{y}_{k-1}}{-g_{k-1}^T d_{k-1}} \tag{6}$$

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

常用的 strong Wolfe-Powell (SWP) 线搜索条件为 $t_k > 0$ 满足以下两式：

$$f(x_k + t_k d_k) - f(x_k) \leqslant \delta t_k g_k^T d_k \tag{7}$$

$$\sigma g_k^T d_k \leqslant g(x_k + t_k d_k)^T d(x_k) \leqslant -\sigma g_k^T d_k \tag{8}$$

其中：$\delta \in (0, 1/2)$，$\sigma \in (\delta, 1)$ 为常数.

共轭梯度法收敛性分析中常用的充分下降性条件为：

$$\exists c > 0,\ \forall k \geqslant 1,\ g_k^T d_k \leqslant -c \| g_k \|^2 \tag{9}$$

① 收稿日期：2007-03-26

基金项目：南宁师范高等专科学校科研资助项目 (2007012).

作者简介：林穗华(1973-)，女，广西龙州人，讲师，主要从事概率统计及最优化理论的研究.

方法(4)、(5)、(6)分别在$\sigma \in (0, 1/4)$、$\sigma \in (0, 1/3)$、$\sigma \in (0, 1/2)$的SWP线搜索下具有充分下降性和全局收敛性[5]. 本文结合方法(4)、(5)、(6)提出带双参数的新共轭梯度法β_k公式如下：

$$\beta_k(\mu_1, \mu_2) = \frac{g_k^T \hat{y}_{k-1}}{-\mu_1 g_{k-1}^T d_{k-1} + \mu_2 d_{k-1}^T y_{k-1} + (1-\mu_1-\mu_2)\| g_{k-1} \|^2} \tag{10}$$

其中参数μ_1, μ_2, $\mu_1+\mu_2 \in [0, 1]$.

显然$\beta_k(0, 0) = \beta_k^{\mathrm{WYL}}$, $\beta_k(0, 1) = \beta_k^{\mathrm{MHS}}$, $\beta_k(1, 0) = \beta_k^{\mathrm{MLS}}$. 对(1)、(2)、(10)的共轭梯度法，当目标函数为二次函数且采用精确线搜索时，有$\beta_k(\mu_1, \mu_2) = \beta_k^{\mathrm{FR}}$.

1 算法及假设

设$\mu = (\mu_1 + \mu_2)$，参数$\theta = \begin{cases} \dfrac{1}{2} & \mu = (0, 0) \\ \dfrac{\mu_1+\mu_2}{2+\mu_2} & \mu \neq (0, 0) \end{cases}$

算法1：

Step 0　给定初值$x_1 \in R^n$, μ_1, $\mu_2 \in [0, 1]$(且满足$\mu_1+\mu_2 \in [0, 1]$)，$\varepsilon \geqslant 0$，$d_1 := -g_1$, $k := 1$. 若$\| g_1 \| \leqslant \varepsilon$，停止.

Step 1　由$\sigma \in (0, \theta)$下的SWP线搜索计算t_k.

Step 2　迭代运算$x_{k+1} := x_k + t_k d_k$, $g_{k+1} := g(x_{k+1})$. 若$\| g_{k+1} \| \leqslant \varepsilon$，停止.

Step 3　由公式(10)计算β_k，由(2)式计算d_{k+1}.

Step 4　$k := k+1$，转Step 1.

假设(ⅰ)设$f(x)$的水平集$\Omega = \{x \in R^n \mid f(x) \leqslant f(x_1)\}$有界，$f(x)$在$\Omega$下方有界.

(ⅱ)设$f(x)$的梯度$g(x)$在Ω上Lipschitz连续，即存在常数$L > 0$，使$\forall x, y \in \Omega$，有

$$\| g(y) - g(x) \| \leqslant L \| y - x \| \tag{11}$$

由假设(ⅰ)知，$\exists B > 0$，使$\| x \| \leqslant B$, $\forall x \in \Omega$.

由假设(ⅱ)知，$\exists \bar{\gamma} > 0$，使$\| g(x) \| \leqslant \bar{\gamma}$, $\forall x \in \Omega$.

2 算法的全局收敛性

引理1　考虑算法1生成的序列$\{g_k\}$和$\{d_k\}$，则存在常数$c > 0$，使得$\forall k \geqslant 1$，(9)式成立；$\forall k \geqslant 2$，有$\beta_k \geqslant 0$.

证　若$\mu = (0, 0)$，因$\beta_k(0, 0) = \beta_k^{\mathrm{WYL}}$，引理结论的证明见文献[6].

若$\mu \neq (0, 0)$，设$c = 1 - \dfrac{2\sigma}{\mu_1 + \mu_2(1-\sigma)}$，由$\sigma \in \left(0, \dfrac{\mu_1+\mu_2}{2+\mu_2}\right)$，知$0 < c < 1$.

$g_1^T d_1 = -\| g_1 \|^2 \leqslant -c \| g_1 \|^2 < 0$，现假设$g_{k-1}^T d_{k-1} \leqslant -c \| g_{k-1} \|^2 < 0$.

由(2)、(8)和(10)式及Cauchy-Schwarz不等式，可得

$$\begin{aligned} \frac{g_k^T d_k}{\| g_k \|^2} &\leqslant -1 + \frac{-2\sigma g_{k-1}^T d_{k-1}}{-[\mu_1 + \mu_2(1-\sigma)] g_{k-1}^T d_{k-1} + (1-\mu_1-\mu_2)\| g_{k-1} \|^2} \\ &\leqslant -1 + \frac{2\sigma g_{k-1}^T d_{k-1}}{-[\mu_1 + \mu_2(1-\sigma)] g_{k-1}^T d_{k-1}} \\ &\leqslant -1 + \frac{2\sigma}{\mu_1 + \mu_2(1-\sigma)} = -c \end{aligned}$$

由数学归纳法可知$\forall k \geqslant 1$，(9)式成立.

由Cauchy-Schwarz不等式可知$g_k^T \hat{y}_{k-1} \geqslant 0$，由(9)式可知$-g_{k-1}^T d_{k-1} > 0$, $d_{k-1}^T y_{k-1} \geqslant -(1-\sigma) g_{k-1}^T d_{k-1} \geqslant (1-\sigma)\| g_{k-1} \|^2 > 0$，所以$\forall k \geqslant 2$, $\beta_k \geqslant 0$. 引理得证.

引理2　设假设(ⅰ)、(ⅱ)成立，考虑算法1生成的序列$\{g_k\}$和$\{d_k\}$，则

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\| d_k \|^2} < +\infty \tag{12}$$

证　由引理 1，$\forall k \geqslant 1$ 有 $g_k^T d_k < 0$.

由(8) 及(11) 式，可得 $-(1-\sigma)g_k^T d_k \leqslant (g_{k+1}-g_k)^T d_k \leqslant L t_k \| d_k \|^2$

由上式得 $t_k \geqslant \dfrac{\sigma-1}{L} \dfrac{g_k^T d_k}{\| d_k \|^2}$

由上式及(7) 式可得 $f_k - f_{k+1} \geqslant -\delta t_k g_k^T d_k \geqslant \delta \dfrac{1-\sigma}{L} \dfrac{(g_k^T d_k)^2}{\| d_k \|^2}$

由假设(i)知$\{f_k\}$单调减且有界，所以$\lim\limits_{k\to\infty} f_{k+1} < +\infty$，上式左、右两边分别求和，得$+\infty > f_1 - \lim\limits_{k\to\infty} f_{k+1} = \sum\limits_{k\geqslant 1}(f_k - f_{k+1}) \geqslant \dfrac{\delta(1-\sigma)}{L}\sum\limits_{k=1}^{\infty}\dfrac{(g_k^T d_k)^2}{\| d_k \|^2}$，(12) 式成立，引理得证.

性质(*)[7] 考虑形如(1) 和(2) 的方法，并假定

$$0 < \gamma \leqslant \| g_k \| \leqslant \bar{\gamma} \tag{13}$$

若存在常数 $b > 1$ 和 $\lambda > 0$，对 $\forall k \geqslant 1$ 有，$|\beta_k| \leqslant b$ 及 $\| s_{k-1} \| \leqslant \lambda \Rightarrow |\beta_k| \leqslant \dfrac{1}{2b}$，其中 $s_{k-1} = x_k - x_{k-1}$，则称方法具有性质(*).

引理 3　设假设(i)、(ii)成立，考虑算法 1 生成的序列$\{\beta_k\}$，假定(13) 成立，则 β_k 满足性质(*).

证　设 $q = c\mu_1 + c\mu_2(1-\sigma) + (1-\mu_1-\mu_2)$，则 $0 < q < 1$. 取 $b = \dfrac{2\bar{r}^2}{qr^2}$，$\lambda = \dfrac{q^2 r^4}{8L\bar{r}^2}$.

因 $1 > q > 0$，$0 < \gamma \leqslant \bar{\gamma}$，故 $b > 1$，$\lambda > 0$.

由式(11)，(8)，(10) 与引理 1 及 Cauchy-Schwarz 不等式可得

$$|\beta_k| = \frac{\| g_k \|^2 \left(1 - \dfrac{g_k^T g_{k-1}}{\| g_k \| \| g_{k-1} \|}\right)}{-\mu_1 g_{k-1}^T d_{k-1} + \mu_2 d_{k-1}^T y_{k-1} + (1-\mu_1-\mu_2)\| g_{k-1} \|^2} \leqslant \frac{2\| g_k \|^2}{q\| g_{k-1} \|^2} \leqslant \frac{2\bar{\gamma}^2}{q\gamma^2} = b$$

设 $\| s_{k-1} \| \leqslant \lambda$，则

$$\begin{aligned}
|\beta_k| &\leqslant \frac{\| g_k \| \left\| g_k - g_{k-1} + g_{k-1} - \dfrac{\| g_k \|}{\| g_{k-1} \|} g_{k-1} \right\|}{q\| g_{k-1} \|^2} \\
&\leqslant \frac{\| g_k \|}{q\| g_{k-1} \|^2}\left(\| g_k - g_{k-1} \| + \big|\, \| g_k \| - \| g_{k-1} \| \,\big|\right) \\
&\leqslant \frac{2\| g_k \| \| g_k - g_{k-1} \|}{q\| g_{k-1} \|^2} \leqslant \frac{2L\| g_k \| \| s_{k-1} \|}{q\| g_{k-1} \|^2} \leqslant \frac{2\bar{r}L\lambda}{qr^2} = \frac{1}{2b}
\end{aligned}$$

引理得证.

引理 4　设假设(i)、(ii)成立，考虑算法 1 生成的序列$\{d_k\}$，若存在常数 $\gamma > 0$，使 $\forall k \geqslant 1$，有 $\| g_k \| \geqslant \gamma$，则$\sum\limits_{k\geqslant 2}\| u_k - u_{k-1} \|^2 < \infty$，其中 $u_k = \dfrac{d_k}{\| d_k \|}$.

设 $\kappa_{k,\Delta}^{\lambda} = \{i \in Z^+ : k \leqslant i \leqslant k+\Delta-1,\ \| s_{i-1} \| > \lambda\}$，$|\kappa_{k,\Delta}^{\lambda}|$ 表示 $\kappa_{k,\Delta}^{\lambda}$ 的元素个数.

引理 5　设假设(i)、(ii)成立，考虑算法 1 生成的序列$\{g_k\}$和$\{s_k\}$. 若存在常数 $\gamma > 0$，使 $\forall k \geqslant 1$，有 $\| g_k \| \geqslant \gamma$，则存在常数 $\lambda > 0$，使得对 $\forall \Delta \in Z^+$ 和指标 k_0，均 $\exists$ 指标 $k \geqslant k_0$，满足 $|\kappa_{k,\Delta}^{\lambda}| > \dfrac{\Delta}{2}$.

定理 1　设假设(i)、(ii)成立，考虑算法 1，则方法给出$\lim\limits_{k\to\infty}\inf \| g_k \| = 0$.

注：由引理 1 知参数 $\beta_k \geqslant 0$，且满足充分下降条件(9). 由引理 3 可知 β_k 具有性质(*). 又由满足 SWP 条件必满足 WWP 条件. 引理 4、引理 5、定理 1 的证明分别与文献[8]中引理 3.3.1、引理 3.3.2、定理 3.3.3 的证明类似. 此处省略具体的证明过程.

3　数值试验

用 Matlab 程序对算法 1 及(3) 式对应的 PRP 方法进行数值试验，其中 $\mu = (\mu_1, \mu_2)$ 为(10) 式对应的参数，试验参数为 $\delta = 0.01$，$\sigma = 0.1$，$\varepsilon = 10^{-5}$，结果见表 1. 表 1 中的其他相关符号意义为：Problem 表示

测试问题的名称；Dim 表示目标函数的维数；NI 表示算法迭代的次数；NF 表示函数值计算的次数；NG 表示函数梯度计算的次数；PRPSWP 表示将算法 1 的 Step3 中用(3) 式代替(10) 式得到的算法.

表 1　PRPSWP 和算法 1 的计算结果

Proplem	Dim	$\mu=(0.9, 0)$ NI/NF/NG	$\mu=(0, 0.2)$ NI/NF/NG	$\mu=(0.7, 0.3)$ NI/NF/NG	$\mu=(0.1, 0.7)$ NI/NF/NG	PRPSWP NI/NF/NG
Rose	2	42/456/110	39/554/104	45/433/117	38/405/98	—
Froth	2	13/31/25	18/90/32	17/88/32	13/77/23	11/76/22
Badscp	2	—	—	—	—	—
Badscb	2	—	—	—	—	—
Beale	2	15/128/28	14/174/25	14/126/26	13/124/25	13/79/24
Jensam	2	10/29/22	10/29/21	10/29/22	10/29/22	10/169/19
Helix	3	38/274/69	42/190/73	54/311/100	67/389/119	35/231/66
Bard	3	23/52/42	24/49/41	22/48/39	25/103/44	30/64/54
Gauss	3	4/8/7	4/8/7	4/8/7	4/8/7	4/57/6
Meyer	3	—	—	—	—	—
Gulf	3	1/2/2	1/2/2	1/2/2	1/2/2	1/2/2
Box	3	—	—	—	—	—
Sing	4	174/395/279	110/461/183	117/475/193	156/454/256	105/367/177
Wood	4	149/1002/281	151/1106/268	187/1097/337	57/337/118	108/544/206
Kowosb	4	59/227/102	62/186/109	95/483/152	92/287/156	112/376/185
Bd	4	—	—	—	—	—
Osb1	5	—	—	—	—	—
Biggs	6	171/553/292	—	—	—	127/512/214
Osb2	11	—	—	—	—	—
Watson	20	1877/4541/2895	2147/5690/3315	1256/3590/1940	2217/5913/3504	2765/6906/4329
Rosex	8	43/510/117	48/359/119	48/391/113	46/594/133	26/472/76
	50	49/410/129	45/586/105	43/597/104	37/479/100	31/633/88
	100	46/573/120	42/603/115	45/266/111	47/547/129	—
Singx	8	51/294/87	46/338/83	43/244/83	69/339/124	206/999/355
Pen1	2	8/41/30	8/41/30	8/41/30	8/41/30	19/308/58
Pen2	4	13/92/31	13/136/31	16/137/36	16/180/37	13/42/30
	50	124/936/263	122/896/252	124/755/257	—	2214/6400/3583
Vardim	2	3/55/6	3/55/6	3/55/6	3/55/6	3/8/7
	50	11/43/36	11/43/36	11/43/36	11/43/36	9/32/25
Trig	3	15/221/28	14/128/28	14/175/26	15/180/29	14/268/26
	50	38/309/67	40/599/63	37/311/66	37/356/66	37/402/60
	100	49/431/90	49/384/87	50/523/85	47/632/87	46/331/88
Bv	3	10/19/16	11/23/19	10/20/16	11/22/18	14/27/24
	10	67/510/105	62/167/108	52/244/90	53/152/90	92/415/154
Ie	3	5/59/9	5/59/9	5/59/9	4/9/8	5/11/10
	50	5/11/10	5/11/10	5/11/10	5/11/10	5/10/9
	100	6/12/12	6/12/12	6/12/12	6/12/12	5/10/10
	200	6/58/9	6/58/9	6/58/9	6/108/10	5/10/10
	500	6/12/12	6/12/12	6/12/12	6/108/10	7/14/14
Trid	3	12/23/21	14/27/26	13/24/22	12/25/23	16/29/25
	50	27/285/40	27/92/43	27/284/39	27/183/36	26/139/41
	100	30/240/42	29/338/42	30/338/42	30/187/38	28/380/35
	200	30/389/44	29/90/39	29/389/44	30/244/46	30/191/42
Band	3	6/12/12	6/12/12	6/12/12	6/12/12	9/62/13
	50	19/77/30	21/79/32	20/78/31	21/79/32	18/131/31
	100	20/78/31	23/80/33	21/79/32	23/81/34	18/187/33
	200	20/78/31	23/80/33	21/79/32	23/80/33	19/184/34
Lin	500	1/3/3	1/3/3	1/3/3	1/3/3	1/3/3
	1000	1/3/3	1/3/3	1/3/3	1/3/3	1/3/3

表 1 中数据显示，对测试问题集的 49 个目标函数，在上述试验参数下，本文提出的(10) 式共轭梯度法

簇在 $\mu=(0.9, 0)$，$(0, 0.2)$，$(0.7, 0.3)$，$(0.1, 0.7)$ 时求解失败的个数分别为 7，8，8，9 个，PRP 方法求解失败的个数为 9 个，适当选取参数 μ 的值本文提出的方法比 PRP 方法取得更好的数值结果.

致谢：本文的完成得到广西大学韦增欣教授的悉心指导，在此对韦增欣教授深表感谢！

参考文献：

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

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

[3] Fletcher R, Reeves C. Function minimization by conjugate gradients [J]. Compute J, 1964, 7(2): 149－154.

[4] Wei Zengxin, Yao Shengwei, Liu Liying. The convergence properties of some new conjugate gradient methods [J]. Applied Mathematics and Computation, 2006, 183(2): 1341－1350.

[5] Yao Shengwei, Wei Zengxin, Huang Hai. A Notes about wyl's conjugate gradient method and its applications [J]. Applied Mathematics and Computation, 2007, 191(2): 381－388.

[6] Huang Hai, Wei Zengxin, Yao Shengwei. 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(2): 1241－1245.

[7] Gilbert J C, Nocedal J. Global convergence properties of conjugate gradient methods for optimization [J]. SIAM Journal of Optimization, 1992, 2(1): 21－42.

[8] 戴彧虹，袁亚湘. 非线性共轭梯度法 [M]. 上海：上海科学技术出版社，2001：30－43.

A Two-parameter Family of Conjugate Gradient Method

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

1. Department of Statistics, East China Normal University, Shanghai 200062, China;

2. Department of Mathematics and Computer Science, Nanning Teachers College, Longzhou, Guangxi 532400, China

Abstract: A new formula of the update β_k, which has two paramerters, is proposed in this paper. The authors prove that, by suitable choosing the parameter σ, under the strong Wolfe-Powell line search condition, the corresponding method has the sufficient descent and global convergence properties. Preliminary numerical results show that the proposed method is very promising.

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

责任编辑　张　栒