广西科学 Guangxi Sciences 2007，14(3):244～246,249

一个基于LS公式修正的新共轭梯度算法*

A New Conjugate Gradient Method for Modified LS Formula

黄　海[1]，林穗华[1]，姚胜伟[2]
HUANG Hai[1]，LIN Sui-hua[1]，YAO Sheng-wei[2]

(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. 广西大学数学与信息科学学院，广西南宁　530004)
(1. Department of Mathematics and Computer Science, Nanning Teachers College, Longzhou, Guangxi, 532400, China; 2. College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, 530004, China)

摘要：基于LS公式 β_k^{LS} 提出一个求解大规模非线性优化问题的新共轭梯度法公式：$\beta_k^* = g_k^T(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1})/(-g_{k-1}^T d_{k-1})$，并证明新公式在 $\sigma \in (0, \frac{1}{2})$ 的SWP线搜索下具有充分下降性和全局收敛性. 新方法的数值试验结果良好.

关键词：无约束优化　共轭梯度法　SWP线搜索　全局收敛

中图法分类号：O224　　**文献标识码**：A　　**文章编号**：1005-9164(2007)03-0244-03

Abstract: A conjugate gradient formula β_k^* based on modified formula β_k^{LS} is proposed. That is $\beta_k^* = g_k^T(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1})/(-g_{k-1}^T d_{k-1})$. It is proved that under the strong Wolfe-Powell line search and the parameter $\sigma \in (0, \frac{1}{2})$, the corresponding method has 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

共轭梯度法具有算法简便，存储需要小等优点，是求解大规模非线性无约束优化问题的一种重要方法. 1991年Liu和Story[1]提出一种共轭梯度法（称为LS共轭梯度法），本文对LS公式进行修正，得到收敛性和数值结果都较好的新共轭梯度算法.

1　新公式

非线性无约束最优化问题的一般形式为 $\min\{f(x) | x \in R^n\}$，其中目标函数 $f: R^n \to R$ 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 $g(x_k)$.

共轭梯度法的迭代公式为：

$$x_{k+1} = x_k + t_k d_k, \tag{1.1}$$

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：

$$d_k = \begin{cases} -g_k, k = 1; \\ -g_k + \beta_k d_{k-1}, k \geqslant 2. \end{cases} \tag{1.2}$$

其中 β_k 为标量参数，一些著名的 β_k 公式[1~5]如下：

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{g_{k-1}^T g_{k-1}}, \tag{1.3}$$

$$\beta_k^{LS} = \frac{g_k^T y_{k-1}}{-g_{k-1}^T d_{k-1}}, \tag{1.4}$$

$$\beta_k^{WYL} = \frac{g_k^T \hat{y}_{k-1}}{g_{k-1}^T g_{k-1}}, \tag{1.5}$$

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

常用的strong Wolfe-Powell (SWP)线搜索条件

收稿日期：2006-12-28

作者简介：黄　海（1969-），男，讲师，主要从事最优化理论与数学建模研究。

*广西自然科学基金项目（0542043）和南宁师范高等专科学校科研项目（2007012）资助。

为，$t_k > 0$ 满足以下两式：

$$f(x_k + t_k d_k) - f(x_k) \leqslant \delta t_k g_k^T d_k, \tag{1.6}$$

$$\sigma g_k^T d_k \leqslant g(x_k + t_k d_k)^T d(x_k) \leqslant -\sigma g_k^T d_k, \tag{1.7}$$

其中参数 $\delta \in (0, \frac{1}{2})$，$\sigma \in (\delta, 1)$ 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：

$$\exists\, c > 0, \forall\, k \geqslant 1, g_k^T d_k \leqslant -c \| g_k \|^2. \tag{1.8}$$

PRP 方法具有很好的数值结果，但即使采用精确线搜索，对一般的非凸函数不必收敛[6]. 由文献[5]首先引进的 $\hat{y}_{k-1}$ 起着非常重要的作用，使得 $\beta_k^{WYL} \geqslant 0$，WYL 方法在 $\sigma \in (0, \frac{1}{4})$ 的 SWP 线搜索下具有充分下降性[7] 和全局收敛性. 基于 β_k^{LS} 公式，考虑 SWP 线搜索，提出新的 β_k 计算公式如下：

$$\beta_k^* = \frac{g_k^T \hat{y}_{k-1}}{-g_{k-1}^T d_{k-1}}. \tag{1.9}$$

2 新算法及假设

算法(＊)

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

步骤 1：由 $\sigma \in (0, \frac{1}{2})$ 的 SWP 线搜索计算 t_k.

步骤 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.

步骤 3：由公式(1.9) 计算 β_k，由(1.2) 式计算 d_{k+1}.

步骤 4：$k := k + 1$，转 Step 1.

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

假设(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

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

由假设(i) 知，存在 $B > 0$，使 $\| x \| \leqslant B, \forall\, x \in \Omega.$ (2.2)

由假设(ii) 知，存在 $\bar{\gamma} > 0$，使 $\| g(x) \| \leqslant \bar{\gamma}, \forall\, x \in \Omega.$ (2.3)

3 算法的充分下降性和全局收敛性

定理 3.1　考虑算法(＊)生成的序列$\{g_k\}$ 和 $\{d_k\}$，则 d_k 具有充分下降性，即存在常数 $c > 0$，使得 $\forall\, k \geqslant 1$，(1.8) 式成立；$\forall\, k \geqslant 2$，有 $\beta_k \geqslant 0$.

证明　设 $c = 1 - 2\sigma$，由 $\sigma \in (0, \frac{1}{2})$，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(1.2) 式和(1.9) 式，可得 $\frac{g_k^T d_k}{\| g_k \|^2} = -1 + (1 - \frac{g_k^T g_{k-1}}{\| g_k \| \| g_{k-1} \|}) \frac{g_k^T d_{k-1}}{-g_{k-1}^T d_{k-1}}$.

由上式、(1.7) 式及 Cauchy-Schwarz 不等式，可得

$$\frac{g_k^T d_k}{\| g_k \|^2} \leqslant -1 + (1 - \frac{g_k^T g_{k-1}}{\| g_k \| \| g_{k-1} \|}) \frac{-\sigma g_{k-1}^T d_{k-1}}{-g_{k-1}^T d_{k-1}} \leqslant -1 + 2\sigma = -c.$$

由数学归纳法可知 $\forall\, k \geqslant 1$，(1.8) 式成立.

由 Cauchy-Schwarz 不等式可知 $g_k^T \hat{y}_{k-1} \geqslant 0$，由(1.8) 式可知 $-g_{k-1}^T d_{k-1} > 0$，所以 $\forall\, k \geqslant 2, \beta_k \geqslant 0$. 定理得证.

引理 3.1　设假设(i)、(ii) 成立，考虑算法(＊)生成的序列$\{g_k\}$ 和$\{d_k\}$，则 Zoutendijk 条件成立，即

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\| d_k \|^2} < +\infty. \tag{3.1}$$

证明　由定理 3.1，$\forall\, k \geqslant 1$ 有 $g_k^T d_k < 0$.

由(1.7) 及(2.1) 式，可得 $-(1-\sigma) g_k^T d_k \leqslant (g_{k+1} - g_k)^T d_k \leqslant L t_k \| d_k \|^2$. (3.2)

由(3.2) 式得 $t_k \geqslant \frac{\sigma - 1}{L} \frac{g_k^T d_k}{\| d_k \|^2}$. (3.3)

由(3.3) 式及(1.8) 式可得 $f_k - f_{k+1} \geqslant -\delta t_k g_k^T d_k \geqslant \delta \frac{1-\sigma}{L} \frac{(g_k^T d_k)^2}{\| d_k \|^2}$.

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 $+\infty > f_1 - \lim\limits_{k \to \infty} f_{k+1} = \sum\limits_{k \geqslant 1} (f_k - f_{k+1}) \geqslant \frac{\delta(1-\sigma)}{L} \sum\limits_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\| d_k \|^2}$，所以(3.1) 式成立，引理得证.

性质(＊)[8]　考虑形如(1.1) 和(1.2) 式的方法，并假定

$$0 < \gamma \leqslant \| g_k \| \leqslant \bar{\gamma}, \tag{3.4}$$

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

引理 3.2　设假设(i)、(ii) 成立，考虑算法(＊)生成的序列$\{\beta_k\}$，假定(3.4) 式成立，则 β_k 满足性质(＊).

证明　取 $b = \frac{2\bar{r}^2}{c r^2}$，$\lambda = \frac{c^2 r^4}{8 L \bar{r}^2}$，因 $1 > c > 0$，$0 < \gamma \leqslant \bar{\gamma}$，故 $b > 1$，$\lambda > 0$.

由(2.1) 式、定理 3.1 及 Cauchy-Schwarz 不等式可得

$$|\beta_k| = \frac{\| g_k \|^2 (1 - \frac{g_k^T g_{k-1}}{\| g_k \| \| g_{k-1} \|})}{-g_{k-1}^T d_{k-1}} \leqslant$$

$$\frac{2\|g_k\|^2}{c\|g_{k-1}\|^2} \leqslant \frac{2\bar{\gamma}^2}{c\gamma^2} = b.$$

设 $\|s_{k-1}\| \leqslant \lambda$,由(1.7)、(1.9)式、定理3.1及Cauchy-Schwarz不等式可得

$$|\beta_k| \leqslant$$

$$\frac{\|g_k\| \left\| g_k - g_{k-1} + g_{k-1} - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right\|}{c\|g_{k-1}\|^2} \leqslant$$

$$\frac{\|g_k\|}{c\|g_{k-1}\|^2}(\|g_k - g_{k-1}\| + |\|g_k\| - \|g_{k-1}\||) \leqslant \frac{2\|g_k\|\|g_k - g_{k-1}\|}{c\|g_{k-1}\|^2} \leqslant$$

$$\frac{2L\|g_k\|\|s_{k-1}\|}{c\|g_{k-1}\|^2} \leqslant \frac{2\bar{\gamma}L\lambda}{c\gamma^2} = \frac{1}{2b}.$$ 引理得证.

引理 3.3 设假设(i)、(ii)成立,考虑算法(*)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$\sum_{k\geqslant 2}\|u_k - u_{k-1}\|^2 < \infty$,其中$u_k = \frac{d_k}{\|d_k\|}$.

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

引理 3.4 设假设(i)、(ii)成立,考虑算法(*)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,均存在指标$k \geqslant k_0$,满足$|k_{k,\Delta}^{\lambda}| > \frac{\Delta}{2}$.

定理 3.2 设假设(i)、(ii)成立,考虑算法(*),则方法给出$\liminf_{k\to\infty}\|g_k\| = 0$.

注 由定理3.1知参数$\beta_k \geqslant 0$,且满足充分下降条件(1.8)式.由引理3.2可知β_k具有性质(*).又由满足SWP条件必满足WWP条件.引理3.3、引理3.4、定理3.2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,在此不再进行具体证明.

4 数值试验

对分别由(1.3)、(1.4)、(1.9)式计算β_k,(1.1)、(1.2)式迭代格式对应的共轭梯度法,我们用Matlab程序进行SWP线搜索条件下的数值试验,试验参数为,$\delta = 0.01$,$\sigma = 0.1$,$\varepsilon = 10^{-5}$,结果见表1.

表1中相关符号表示如下:PRPSWP:(1.3)+SWP;LSSWP:(1.4)+SWP;MLSSWP:(1.9)+SWP;Problem:测试问题的名称;Dim:目标函数的维数;NI/NF/NG:迭代次数/目标函数计算次数/目标梯度函数计算次数.

表 1 PRPSWP、LSSWP 和 MLSSWP 的计算结果

Table 1 The results for PRPSWP,LSSWP and MLSSWP

Proplem	Dim	NI/NF/NG		
		PRPSWP	LSSWP	MLSSWP
Rose	2	—	—	54/732/143
Froth	2	11/76/22	9/68/17	15/36/29
Badscp	2	—	—	—
Badscb	2	—	—	—
Beale	2	13/79/24	12/28/23	14/126/26
Jensam	2	10/169/19	8/116/15	10/29/22
Helix	3	35/231/66	113/383/192	37/225/69
Bard	3	30/64/54	48/197/87	84/218/138
Gauss	3	4/57/6	3/55/4	4/8/7
Meyer	3	—	—	—
Gulf	3	1/2/2	1/2/2	1/2/2
Box	3	—	—	—
Sing	4	105/367/177	314/727/493	305/856/486
Wood	4	108/544/206	65/450/135	200/996/362
Kowosb	4	112/376/185	59/366/94	118/672/193
Bd	4	—	—	—
Osb1	5	—	—	—
Biggs	6	127/512/214	—	52/165/92
Osb2	11	—	—	320/1337/531
Watson	20	2765/6906/4329	2853/7523/4479	1372/3201/2141
Rosex	8	26/472/76	—	46/415/122
	50	31/633/88	—	35/420/100
	100	—	23/367/71	43/565/111
Singx	8	206/999/355	34/176/65	225/735/359
Pen1	2	19/308/58	23/210/84	8/41/30
Pen2	4	13/42/30	12/81/30	16/137/35
	50	2214/6400/3583	390/2156/719	225/1136/408
Vardim	2	3/8/7	3/7/7	3/55/6
	50	9/32/25	10/91/35	11/43/36
Trig	3	14/268/26	12/217/23	17/133/32
	50	37/402/60	39/220/72	40/314/72
	100	46/331/88	46/616/82	49/377/84
Bv	3	14/27/24	8/16/13	11/119/17
	10	92/415/154	52/197/86	73/279/119
Ie	3	5/11/10	5/11/10	5/59/9
	50	5/10/9	5/10/9	5/11/10
	100	5/10/10	5/10/10	6/12/12
	200	5/10/10	5/10/10	6/58/9
	500	7/14/14	7/14/14	6/12/12
Trid	3	16/29/25	12/24/21	13/26/24
	50	26/139/41	26/331/37	27/237/41
	100	28/380/35	28/185/36	30/336/40
	200	30/191/42	29/337/41	30/437/43
Band	3	9/62/13	—	6/12/12
	50	18/131/31	19/84/32	19/77/30
	100	18/187/33	19/178/32	20/78/31
	200	19/184/34	19/281/35	20/78/31
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

(下转第 249 页 Continue on page 249)

将主观权重与客观权重加权综合，其加权系数由数学规划模型求出.该方法得出的结果较好地反映了主观程度和客观程度，并通过算例说明该方法是有效可行的.

参考文献：

[1] 樊治平，潘德惠.多属性决策的一种主客观综合法[J].系统工程，1995，13(5)：28-31.

[2] MA JIAN，FAN ZHIPING，HUANG LIHUA. A subjective and objective integrated approach to determine attribute weights[J]. European Journal of Operational Research，1999，112：397-404.

[3] 徐泽水，达庆利.多属性决策的组合赋权方法研究[J].中国管理科学，2002，10(2)：84-87.

[4] 张海涛，刘超英，田水.权重确定的主客观综合法[J].江汉大学学报：自然科学版，2004，32(4)：63-65.

[5] 任全，李为民，张文.客观赋权法指导下的部分权重信息多属性决策方法研究[J].数学的实践与认识，2004，34(7)：86-90.

[6] 谭旭，陈英武，高妍方.一种新的基于组合赋权的区间型多属性决策方法[J].系统工程，2006，24(4)：111-114.

[7] 王中兴，李桥兴.依据主、客观权重集成最终权重的一种方法[J].应用数学与计算数学学报，2006，20(1)：87-92.

[8] 赵萱，张全，樊治平.多属性决策中权重确定的主客观赋权法[J].沈阳工业大学学报，1997，19(4)：95-98.

[9] 许叶军，达庆利.基于理想点的多属性决策主客观赋权法[J].工业工程与管理，2005，10(4)：45-47.

[10] 魏巍贤，冯佳.多目标权系数的组合赋值方法研究[J].系统工程与电子技术，1998，20(2)：14-16.

[11] 陈华友.多属性决策中的一种最优组合赋权方法研究[J].运筹与管理，2003，12(2)：6-10.

[12] 徐泽水.权重信息完全未知且对方案有偏好的多属性决策方法[J].系统工程理论与实践，2003，12：100-103.

（责任编辑：尹　闯　邓大玉）

（上接第 246 页 Continue from page 246）

表 1 中数据显示，对测试问题集中的 53 个目标函数，在试验参数为 $\delta=0.01,\sigma=0.1,\varepsilon=10^{-5}$ 的情况下，LSSWP、PRPSWP、MLSSWP 方法求解失败的个数分别为 12、9、6 个.

致谢：

本文系作者在广西大学访学期间完成，得到韦增欣教授的关怀指导，对本文的完成提出宝贵的意见，我们在此对韦增欣教授深表感谢！

参考文献：

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

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

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

[4] LIU Y，STORY C. Efficient generalized conjugate gradient algorithms Part 2：Implementation[J]. JOTA，1991，69(1)：139-152.

[5] WEI Z X，YAO S W，LIU L Y. The convergence properties of some new conjugate gradient methods[J]. Applied Mathematics and Computation，2006，183(2)：1341-1350.

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

[7] 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.

[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，2001：38-43.

（责任编辑：邓大玉）