广西科学 Guangxi Sciences 2007，14(1)：44～46

一类 Wolfe 搜索下的共轭梯度法及其全局收敛性*

Global Convergence for a Class of Conjugate Gradient Method with Wolfe Search

董晓亮，李郴良，唐清干
DONG Xiao-liang，LI Chen-liang，TANG Qing-gan

（桂林电子科技大学计算科学与数学系，广西桂林 541004）
（Department of Computing Science and Mathematics，Guilin University of Electronic Technology，Guilin，Guangxi，541004，China）

摘要：构造一个新的求解无约束优化问题的非线性共轭梯度公式中 β_k 参数的计算公式，采用该公式得到 Wolfe 非精确线搜索的新算法，并证明新算法具有全局收敛性，并用数值实验验证新算法是有效的.

关键词：无约束优化　共轭梯度法　全局收敛性　Wolfe 准则
中图法分类号：O224　　**文献标识码**：A　　**文章编号**：1005-9164(2007)01-0044-03

Abstract：A new nonlinear conjugate gradient method is proposed. A formula is constructed to compute the parameter β_k of the method. The algorithm for an inexact line search of Wolfe conditions is obtained by this method. Global convergence of the algorithm is proved. Numerical tests show that the algorithm is effective.

Key words：unconstrained optimization，conjugate gradient method，global convergence，Wolfe condition

共轭梯度法是求解无约束优化问题 $\min f(x)$ 的一类重要而有效的方法，其主要迭代公式如下：

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

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

其中 $g_k = \nabla f_k$，d_k 是搜索方向，α_k 是步长因子，β_k 是参数. 自 1964 年 Fletcher 和 Reeves 首先提出非线性共轭梯度法以来，参数 β_k 的取法一直备受关注. 其中，$\beta_k^{HS} = \dfrac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}$（HS 公式[1]）数值效能表现良好，特别适于大规模无约束最优化问题[2]. 对凸二次函数在精确线搜索下它有有限收敛性，但若目标函数为非严格凸二次函数，即使在精确线搜索下也未必有有限收敛性，全局收敛性也不能得到保证[3]. 近来，文献[4～8] 对 HS 公式就公式的修正以及全局收敛性等方面做了新探索. 本文在上述文献的基础上，结合 HS 公式的特点和优点构造出一个新的 β_k 的取法：

$$\beta_k^{\theta} = \frac{\| g_{k+1} - g_k \|^2}{d_k^T(\theta g_{k+1} - g_k)} (\theta\sigma < 1). \tag{3}$$

并对基于(3) 式得到的新算法的全局收敛性进行分析，同时验证了新算法的有效性.

1　预备知识

在构造共轭梯度法时，通常步长 α_k 满足所谓的 Wolfe 条件：

$$f(x_k) - f(x_k + \alpha_k d_k) \geqslant -\rho\alpha_k g_k^T d_k, \tag{4}$$

$$\sigma g_k^T d_k \leqslant g(x_k + \alpha_k d_k)^T d_k \leqslant 0, \tag{5}$$

其中，$0 < \rho < \sigma < 1$.

为方便计，本文约定 $q_k = \dfrac{\| g_{k+1} - g_k \|^2}{\| g_{k+1} \|^2}$. 另外，为证明需要，特做以下必要的假设：

假设 1　$f(x)$ 在水平集 $L = \{x \in R^n | f(x) \leqslant f(x_0)\}$ 有界；

假设 2　$f(x)$ 在水平集的某一邻域 D 内连续可微，$g(x)$ 满足 Lispchitz 条件，即存在 $M > 0$，使

$$\| g(x) - g(y) \| \leqslant M \| x - y \|, x, y \in D; \tag{6}$$

假设 3　$f(x)$ 在水平集上是一致凸函数，即存在常数 $\lambda > 0$，使对任意 $x, y \in D$，有 $(x - y)^T[g(x)$

收稿日期：2006-03-28
修回日期：2006-06-20
作者简介：董晓亮（1981-），男，硕士研究生，主要从事微分差分方程的数值解、变分不等式等研究。
* 国家自然科学基金（10371035）资助项目。

$-g(y)]\geqslant\lambda\|x-y\|^2$.

引理1[8] 若假设3成立，$f(x_k+\alpha_k d_k)$ 下有界，则对满足(4)式和(5)式的 $\alpha_k>0$ 有

$$f(x_k)-f(x_{k+1})\geqslant m\|\alpha_k d_k\|^2. \tag{7}$$

引理2 若假设1和假设3成立，α_k 满足(4)式和(5)式，则当 $k\to\infty$ 时，$\|x_{k+1}-x_k\|\to 0$.

证明 由假设1中 $f(x)$ 有界和(7)式得

$$+\infty>\frac{1}{m}\sum_{k=1}^{\infty}[f(x_k)-f(x_{k+1})]\geqslant \sum_{k=1}^{\infty}\|\alpha_k d_k\|^2=\sum_{k=1}^{\infty}\|x_{k+1}-x_k\|, \tag{8}$$

则级数 $\sum_{k=1}^{\infty}\|x_{k+1}-x_k\|$ 收敛，也即当 $k\to\infty$ 时，$\|x_{k+1}-x_k\|\to 0$.

引理3 若 β 取 β_k^θ 时，目标函数满足假设1和假设2，且 α_k 满足Wolfe条件(4)式和(5)式，其中，$0<\rho<\sigma<1$ 且 $\theta\sigma<1$，则当 $\|g_k\|\neq 0$，有

$$1\leqslant -d_k^T g_k/\|g_k\|^2\leqslant 1+S\sigma q_{k-1}, \tag{9}$$

其中，$S=\max(1,(1-\theta\sigma)^{-1})$，从而 d_k 是下降方向.

证明 首先考虑 $\theta<0$ 时的情形. 因为 $-g_1^T d_1=\|g_1\|^2$，则当 $k=1$ 结论是成立的. 若 $k\geqslant 1$ 时结论成立，且 $g_{k+1}\neq 0$，则考虑 $k+1$ 的情形，由 d_{k-1} 的定义得

$$\frac{-g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2}=1-\beta_k^\theta\frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2}. \tag{10}$$

另外，注意在由(5)式的条件下有

$$0\leqslant\theta g_{k+1}^T d_k\leqslant\theta\sigma g_k^T d_k, \tag{11}$$

同时 $-g_k^T d_k\geqslant 0$，因此可得

$$0<\frac{\|g_{k-1}-g_k\|^2}{(1-\theta\sigma)(-g_k^T d_k)}\leqslant\beta_k^\theta\leqslant\frac{\|g_{k+1}-g_k\|^2}{-g_k^T d_k}, \tag{12}$$

$$0\leqslant-\beta_k^\theta\frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2}\leqslant\sigma\frac{\|g_{k+1}-g_k\|^2}{\|g_{k+1}\|^2}, \tag{13}$$

则从(10)和(13)式可得到 $k+1$ 时的结论成立.

若 $0<\theta<1/\sigma$，d_k 仍然是一个下降方向. 事实上，我们可以推出与 $\theta<0$ 对应的式子如下：

$$0<\frac{\|g_{k+1}-g_k\|^2}{(g_k^T d_k)}\leqslant\beta_k^\theta\leqslant\frac{\|g_{k+1}-g_k\|^2}{(-g_k^T d_k)(1-\theta\sigma)}, \tag{14}$$

$$0\leqslant-\beta_k^\theta\frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2}\leqslant\frac{\sigma}{1-\theta\sigma}\frac{\|g_{k+1}-g_k\|^2}{\|g_{k+1}\|^2}, \tag{15}$$

$$1\leqslant-\frac{d_k^T g_k}{\|g_k\|^2}\leqslant 1+\frac{\sigma q_{k-1}}{1-\theta\sigma}. \tag{16}$$

其过程类似上述证明过程，此处不再赘述.

引理4[8] 若假设1和假设2成立，搜索方向 d_k 满足 $g_k^T d_k<0$，且步长 α_k 满足(4)式和(5)式，则

$$\sum_{k\geqslant 1,\|d_k\|\neq 0}\frac{(d_k^T g_k)^2}{\|d_k\|^2}<+\infty. \tag{17}$$

2 算法的步骤

步骤1：选定初始步，给出 $x_1\in R^n$，置 $k=1$，$a=0,b=10^4,\alpha=1,\theta=\theta_0$ 及收敛精度 $\varepsilon>0$.

步骤2：计算 $g_k=g(x_k)$. 若 $\|g_k\|<\varepsilon$，则停止计算，输出 x_k；否则令 $d_k=-g_k+\beta_{k-1}d_{k-1}$，$\beta_{k-1}=\begin{cases}0,k=1;\\ \beta_{k-1}^\theta,k>1.\end{cases}$

步骤3：置 ρ,σ_1 满足 $0<\rho<\sigma<1$ 和 $\theta\sigma<1$，计算 $f_{k+1}=f(x_k+\alpha d_k)$，$g_{k+1}=g(x_k+\alpha d_k)$. 若 α 满足(4)式和(5)式，则令 $\alpha_k=\alpha$，转步骤6；否则，若 α 不满足(4)式，则转步骤4；若 α 不满足(5)式，则转步骤5.

步骤4：令 $b=\alpha,\alpha=\dfrac{a+\alpha}{2}$，转步骤3.

步骤5：令 $a=\alpha,\alpha=\min(2\alpha,\dfrac{a+\alpha}{2})$，转步骤3.

步骤6：令 $x_{k+1}=x_k+\alpha_k d_k$，置 $k=k+1$，转步骤2.

3 全局收敛性分析

定理1 假设1，假设2和假设3成立，$\{x_k\}$ 由迭代公式(1)和(2)产生，$\beta_k=\beta_k^\theta$，α_k 满足(4)式和(5)式，其中的 $0<\rho<\sigma<1$，则必在有限步内在 x_k 处有 $g_k=0$ 或者 $\lim\limits_{k\to\infty}\|g_k\|=0$ 成立.

证明 不失一般性，我们假定对于任何 k，均有 $g_k\neq 0$. 可分别由(9)式和(12)～(15)式得：

$$0<\beta_k^\theta\leqslant S\frac{\|g_{k+1}-g_k\|^2}{\|g_k\|^2}, \tag{18}$$

$$0\leqslant-\beta_k^\theta g_{k+1}^T d_k\leqslant S\sigma\|g_{k+1}-g_k\|^2. \tag{19}$$

又由(18)式和(19)式得：

$$\|d_{k+1}\|^2=\|g_{k+1}\|^2+(\beta_k^\theta)^2\|d_k\|^2-2\beta_k^\theta g_{k+1}^T d_k\leqslant\|g_{k+1}\|^2+S^2\frac{\|g_{k+1}-g_k\|^4}{\|g_k\|^4}\|d_k\|^2+2S\sigma\|g_{k+1}-g_k\|^2, \tag{20}$$

若令 $t_k=\dfrac{\|d_k\|^2}{\|g_k\|^4}$，并在(20)式两边同时除以 $\|g_{k+1}\|^4$，则

$$t_{k+1}\leqslant\|g_{k+1}\|^{-2}+S^2\frac{\|g_{k+1}-g_k\|^4}{\|g_k\|^4\|g_{k+1}\|^4}\|d_k\|^2+2S\sigma\frac{\|g_{k+1}-g_k\|^2}{\|g_{k+1}\|^4}\leqslant\|g_{k+1}\|^{-2}+S^2\frac{\|g_{k+1}-g_k\|^4}{\|g_{k+1}\|^4}t_k+2S\sigma\frac{\|g_{k+1}-g_k\|^2}{\|g_{k+1}\|^4}.$$

若定理不成立，反设 $\lim\limits_{k\to\infty}\|g_k\|\neq 0$，则存在 $\varepsilon>0$，当 k 充分大时，有 $\|g_k\|>\varepsilon$. 同时，由(6)式和引理2得 $q_k\to 0$. 故不妨设 $0<q_k<1$，又由 S 的取法知其

有界，则当 k 充分大的时候，必有 $0 < q_k S < 1$. 设 $N_k = \varepsilon^{-2} + 2S\sigma \frac{\| g_{k+1} - g_k \|^2}{\| g_{k+1} \|^4} > 0$，则有：

$$t_{k+1} \leqslant t_k + N_k,$$

$$\infty > \sum_{k \geqslant 1} \frac{(g_k^T d_k)^2}{\| d_k \|^2} = \sum_{k \geqslant 1} \frac{(g_k^T d_k)^2 / \| g_k \|^4}{t_k} \geqslant$$

$$\sum_{k \geqslant 1} \frac{1}{t_1 + N_k(k-1)} = \infty. \tag{21}$$

由(21)式矛盾可推知定理成立.

4 数值实验

例 1 应用本文算法求解极小化函数 $f(x) = (x_1 - 2)^4 + (x_1 - 2)^2 x_2^2 + (x_2 + 1)^2$，选取初始点 $x_0 = [1,1]$，经过 8 次迭代得到最优点 $x^* = [2, -1]$.

从表 1 中可以看出，本算法较文献[9]中的

表 1 数值试验结果

Table 1 Numerical experiment results of algorithms

算法 Algorithm	x_k	$f(x_k)$
BFGS 算法[9] BFGS algorithm	$x_0 = (1,1)$	6.0
	$x_1 = (1, -0.5)$	1.5
	$x_2 = (1.45, -0.3875)$	5.12×10^{-1}
	$x_3 = (1.58889, -0.63729)$	2.29×10^{-1}
	$x_4 = (1.82541, -0.97156)$	3.05×10^{-2}
	$x_5 = (1.94063, -1.07056)$	8.33×10^{-3}
	$x_6 = (1.96414, -1.04509)$	3.44×10^{-3}
	$x_7 = (1.99521, -1.00171)$	2.59×10^{-5}
	$x_8 = (2.00007, -1.00043)$	1.89×10^{-7}
	$x_9 = (1.99998, -0.99995)$	2.52×10^{-9}
	$x_{10} = (2.0, -1.0)$	0
本文算法 New algorithm	$x_0 = (1,1)$	6.0
	$x_1 = (2.5, -0.5)$	0.375
	$x_2 = (2.07813, -0.82813)$	0.004782
	$x_3 = (1.97052, -0.93661)$	0.001612
	$x_4 = (1.96625, -1.0206)$	8.438×10^{-5}
	$x_5 = (2.005113, -1.00761)$	8.127×10^{-7}
	$x_6 = (2.000732, -0.999473)$	4.2895×10^{-10}
	$x_7 = (1.99999, -0.99998)$	2.842×10^{-14}
	$x_8 = (2,1)$	0

BFGS 方法迭代略快，从而验证了算法的可行性和有效性.

参考文献：

[1] HESTENES M R, STIEFEL E L. Methods of conjugate gradients for solving linear systems[J]. J Res Nat Bur Standards Sect, 1952, 49: 409-436.

[2] AL-BAALI M. Descent property and global convergence of the Fletcher-Reeves method with inexact line search [J]. IMA J Numer Anal, 1985, 5: 121-124.

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

[4] 戴志锋，陈兰平. 一种混合的 HS－DY 共轭梯度法[J]. 计算数学，2005，27(4)：429-436.

[5] 戚后铎，韩继业，刘光辉. 修正 Hestenes-Stiefel 共轭梯度法[J]. 数学年刊，1996，17A(3)：277-284.

[6] 时贞军. 改进 HS 共轭梯度法及其全局收敛性[J]. 计算数学，2001，23(4)：393-406.

[7] 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，2001：30-50.

[8] 袁亚湘，孙文瑜. 最优化理论与方法[M]. 北京：科学出版社，1997：60-68.

[9] 孙文瑜，徐成贤，朱德通. 最优化方法[M]. 北京：高等教育出版社，2004：134.

(责任编辑：韦廷宗)

《广西科学》杂志被评为中国精品科技期刊

广西科学院、广西科学技术协会主办，广西科学技术厅、广西教育厅协办的《广西科学》杂志，刊登的论文水平高，较好地反映了广西科学技术的水平和发展；刊物的质量高，2005 年被广西出版局列在广西差错率最少的科技刊物的第一位；它在自治区内外的影响大，越来越受到广大科技人员和其他各方面的关注和重视，论文引用率逐年增加。最近，《广西科学》杂志被中国科协评为 2006 年中国精品科技期刊，获得中国科协精品科技期刊工程项目(C 类)资助。这是广西唯一入选的刊物。

(罗海鹏)