DOI:10.13656/j.cnki.gxkx.2012.02.016

广西科学 Guangxi Sciences 2012,19(2):104～107

一种新的修正共轭梯度算法*

A New Modified Conjugate Gradient Method

赵许培，杨英芝，袁功林

ZHAO Xu-pei，YANG Ying-zhi，YUAN Gong-lin

（广西大学数学与信息科学学院，广西南宁 530004）

（College of Mathematics and Information Science，Guangxi University，Nanning，Guangxi，530004，China）

摘要：给出求解无约束优化问题的一个新的共轭梯度算法，证明该算法在强 Wolfe 线搜索下具有全局收敛性和良好的数值表现.

关键词：无约束最优化　共轭梯度算法　线搜索　全局收敛性

中图法分类号：O211　**文献标识码：**A　**文章编号：**1005-9164(2012)02-0104-04

Abstract：A new conjugate gradient method for solving unconstrained optimization problem is given. The global convergence of the method is established. Numerical results show that the method has good performance.

Key words：unconstrained optimization，conjugate gradient method，line research，global convergence

对于无约束最优化问题

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

其中 $f(x): R^n \to R$，是连续可微的函数，g 为梯度向量。求解问题(0.1)的共轭梯度法的迭代公式为 $x_{k+1} = x_k + \alpha_k d_k$. 其中 α_k 为某种确定的线搜索给出的步长，搜索方向 d_k 定义为

$$d_k = \begin{cases} -g_0, \text{if } k = 0, \\ -g_k + \beta_k d_{k-1}, \text{if } k \geqslant 1. \end{cases}$$

β_k 为参数。关于参数 β_k 的不同取法就对应不同的共轭梯度法，著名的梯度法[1～6]有

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}},$$

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

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

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

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

其中 $y_{k-1} = g_k - g_{k-1}$，$\|\cdot\|$ 为欧几里德范数. 已有许多学者在共轭梯度法方面做了进一步的工作，取得了丰硕的成果[7～11]. 目前，在共轭梯度算法领域，如何寻找一个新的关于参数 β_k 的公式，使得公式既能全局收敛又有好的数值结果依然是项十分有意义的工作. 本文给出了一种新的共轭梯度算法，该算法具有充分下降性，全局收敛性和理想的数值验证结果.

1　新的共轭梯度算法

文献[12]在文献[7]的基础上对PRP算法进行了修正，该公式为

$$d_k = \begin{cases} -g_k, \text{if } k = 0, \\ -\left(1 + \beta_k \frac{d_{k-1}^T g_k}{\| g_k \|^2}\right) g_k + \beta_k d_{k-1}, \text{if } k \geqslant 1. \end{cases} \tag{1.1}$$

其中

$$\beta_k = \frac{g_k^T y_{k-1}}{\| g_{k-1} \|^2} - \min\left\{ \frac{g_k^T y_{k-1}}{\| g_{k-1} \|^2}, u \frac{\| y_{k-1} \|^2 g_k^T d_{k-1}}{\| g_{k-1} \|^4} \right\},$$

收稿日期：2011-08-29

修回日期：2011-10-05

作者简介：赵许培(1986-)，男，硕士，主要从事最优化理论及应用研究。

* 国家自然科学基金项目(11161003)，广西自然科学基金项目(2012GXNSFAA053002)，广西高校优秀人才资助计划项目，广西教育厅项目(201012MS013)和广西大学科研基金项目(XJZ110632)资助。

而且 $u > \frac{1}{4}$. 由于此公式只含有梯度信息，而在拟牛顿方法中，有一些新公式不但含有梯度信息，还含有函数信息. 例如，文献[3]中的 y_{k-1}^{*}，其表达式为

$$y_{k-1}^{*} = y_{k-1} + \frac{\max\{\rho_{k-1}, 0\}}{\| s_{k-1} \|^{2}} s_{k-1},$$

其中

$$\rho_{k-1} = 2[f(x_{k-1}) - f(x_k)] + (g(x_k) + g(x_{k-1}))^{T} s_{k-1}, s_{k-1} = \alpha_{k-1} d_{k-1}.$$

受此启发，我们用文献[3]中的 y_{k-1}^{*} 来替换(1.1)式中的 y_{k-1}.

通过以上分析得到搜索方向 d_k 为

$$d_k = \begin{cases} -g_k, \text{if } k = 0, \\ -\left(1 + \beta_k^{*} \frac{d_{k-1}^{T} g_k}{\| g_k \|^{2}}\right) g_k + \beta_k^{*} d_{k-1}, \text{if } k \geqslant 1. \end{cases} \quad (1.2)$$

其中

$$\beta_k^{*} = \frac{g_k^{T} y_{k-1}^{*}}{\| g_{k-1} \|^{2}} - \min\left\{ \frac{g_k^{T} y_{k-1}^{*}}{\| g_{k-1} \|^{2}}, u \frac{\| y_{k-1}^{*} \|^{2} g_k^{T} d_{k-1}}{\| g_{k-1} \|^{4}} \right\}. \quad (1.3)$$

当 $\frac{g_k^{T} y_{k-1}^{*}}{\| g_{k-1} \|^{2}} \leqslant u \frac{\| y_{k-1}^{*} \|^{2} g_k^{T} d_{k-1}}{\| g_{k-1} \|^{4}}$ 时，$\beta_k^{*} = 0$.

当 $\frac{g_k^{T} y_{k-1}^{*}}{\| g_{k-1} \|^{2}} > u \frac{\| y_{k-1}^{*} \|^{2} g_k^{T} d_{k-1}}{\| g_{k-1} \|^{4}}$ 时，$\beta_k^{*} = \frac{g_k^{T} y_{k-1}^{*}}{\| g_{k-1} \|^{2}} - u \frac{\| y_{k-1}^{*} \|^{2} g_k^{T} d_{k-1}}{\| g_{k-1} \|^{4}}$.

由于 $\beta_k^{*} = 0$ 时较为简单，故省略. 这里主要研究当 $\beta_k^{*} = \frac{g_k^{T} y_{k-1}^{*}}{\| g_{k-1} \|^{2}} - u \frac{\| y_{k-1}^{*} \|^{2} g_k^{T} d_{k-1}}{\| g_{k-1} \|^{4}}$ 时，算法的全局收敛性质.

算法 1

步骤 1 选择一个初始点 $x_0 \in R^n$，$\varepsilon \geqslant 0$；令 $d_0 = -g_0$，$k = 0$；

步骤 2 利用(1.3)式计算 β_k^{*}，利用(1.2)式计算 d_k；

步骤 3 计算 g_k，如果 $\| g_k \| \leqslant \varepsilon$，则终止；否则利用强 Wolfe 非精确线搜索

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \delta \alpha_k g_k^{T} d_k \quad (1.4)$$

和

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

其中 $0 < \delta < \sigma < 1$，$\delta < \frac{1}{2}$，求 α_k；

步骤 4 令 $x_{k+1} = x_k + \alpha_k d_k$；

步骤 5 令 $k = k + 1$，转步骤 3 计算.

2 算法的全局收敛性

讨论各种共轭梯度算法在不同线搜索下的收敛性质，即要证明

$$\liminf_{k \to \infty} \| g_k \| = 0. \quad (2.1)$$

常用的方法是反证法. 即假定(2.1)式不成立，那么存在常数 $\gamma > 0$，使得 $\| g_k \| > \gamma$，$\forall k \geqslant 1$. 然后再通过证明

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

而导出了与 Zoutendijk 条件矛盾.

假设 1 水平集 $\Omega = \{x \in R^n : f(x) \leqslant f(x_0)\}$ 为有界集.

假设 2 在 Ω 的某个邻域 N，目标函数 $f(x)$ 连续可微，且其梯度 $g(x)$ 满足 Lipschitz 条件，即存在一个常数 $L > 0$，使得

$$\| g(x) - g(y) \| \leqslant L \| x - y \|, \forall x, y \in N. \quad (2.2)$$

由假设 1，即存在一个正常数 B，使得

$$\| x - y \| \leqslant B, \forall x, y \in N,$$

那么存在 $\gamma > 0$，使得

$$\| g(x) \| \leqslant \gamma, \forall x, \in N.$$

假设 3 目标函数 $f(x)$ 为一致凸函数，即存在一个正常数 m，使得

$$m \| d \|^{2} \leqslant d^{T} \nabla^{2} f(x) d, \forall x, d \in R^n, \quad (2.3)$$

其中 $\nabla^{2} f(x)$ 为 $f(x)$ 在 x 处的 Hesse 矩阵.

引理 1 对于(1.2)式，有

$$g_k^{T} d_k = -\| g_k \|^{2}. \quad (2.4)$$

证明 由(1.2)式可知，当 $k = 0$ 时，

$$g_0^{T} d_0 = -\| g_0 \|^{2}.$$

当 $k \geqslant 1$ 时，

$$g_k^{T} d_k = -\left(1 + \beta_k^{*} \frac{d_{k-1}^{T} g_k}{\| g_k \|^{2}}\right) \| g_k \|^{2} + \beta_k^{*} g_k^{T} d_{k-1} = -\| g_k \|^{2} - \beta_k^{*} g_k^{T} d_{k-1} + \beta_k^{*} g_k^{T} d_{k-1} = -\| g_k \|^{2}.$$

故对于所有的 $k \geqslant 0$ 均有(2.4)式成立，证明完毕.

引理 2 如果假设 3 成立，则有

$$q_1 \alpha_k \| d_k \|^{2} \leqslant -g_k^{T} d_k, \quad (2.5)$$

其中 $q_1 = \frac{1}{2}(1 - \delta)^{-1} m$.

证明 利用泰勒展式，有

$$f(x_{k+1}) = f(x_k) + g_k^{T} s_k + \frac{1}{2} s_k G_k s_k,$$

其中 $s_k = x_{k+1} - x_k = \alpha_k d_k$，$G_k = \nabla^{2} f(x_k + t s_k)$，$0 < t < 1$. 由(1.4)式可得

$$f(x_{k+1}) - f(x_k) \leqslant \delta g_k^{T} s_k,$$

$$g_k^{T} s_k + \frac{1}{2} s_k G_k s_k \leqslant \delta g_k^{T} s_k,$$

$$\frac{1}{2}s_k G_k s_k \leqslant (1-\delta)(-g_k^T s_k).$$

由(2.3)式可得

$$\frac{1}{2}(1-\delta)^{-1} m\alpha_k \| d_k \|^2 \leqslant - g_k^T d_k.$$

令 $q_1 = \frac{1}{2}(1-\delta)^{-1} m$，则有 $q_1 \alpha_k \| d_k \|^2 \leqslant -g_k^T d_k$，证明完毕.

引理3 若假设1，假设2和假设3都成立，则有

$$\| d_k \| \leqslant q_2 \| g_k \|, \tag{2.6}$$

其中 $q_2 = 1 + \frac{4L}{q_1} + u\frac{8L^2}{q_1}$.

证明 因为

$$\beta_k^* = \frac{g_k^T y_{k-1}^*}{\| g_{k-1} \|^2} - u\frac{\| y_{k-1}^* \|^2 g_k^T d_{k-1}}{\| g_{k-1} \|^4},$$

由(2.4)式可得

$$\beta_k^* = \frac{g_k^T y_{k-1}^*}{-d_{k-1}^T g_{k-1}} - u\frac{\| y_{k-1}^* \|^2 g_k^T d_{k-1}}{(-d_{k-1}^T g_{k-1})^2},$$

$$| \beta_k^* | \leqslant \frac{| g_k^T y_{k-1}^* |}{| -d_{k-1}^T g_{k-1} |} + u\frac{\| y_{k-1}^* \|^2 | g_k^T d_{k-1} |}{| -d_{k-1}^T g_{k-1} |^2}.$$

再由文献[13]得到

$$\| y_{k-1}^* \| \leqslant 2\| y_{k-1} \|,$$

则有

$$| \beta_k^* | \leqslant \frac{2\| g_k \| \| y_{k-1} \|}{| -d_{k-1}{}^T g_{k-1} |} + u\frac{4\| y_{k-1} \|^2 | g_k^T d_{k-1} |}{| -d_{k-1}^T g_{k-1} |^2}.$$

由(2.2)式和(2.5)式得到

$$| \beta_k^* | \leqslant \frac{2L\alpha_{k-1} \| d_{k-1} \| \| g_k \|}{q_1 \alpha_{k-1} \| d_{k-1} \|^2} + u\frac{4L^2 \alpha_{k-1}^2 \| d_{k-1} \|^2}{q_1^2 \alpha_{k-1}^2 \| d_{k-1} \|^4} \| g_k \| \| d_{k-1} \| = \frac{2L\| g_k \|}{q_1 \| d_{k-1} \|} + u\frac{4L^2 \| g_k \|}{q_1^2 \| d_{k-1} \|} = \left(\frac{2L}{q_1} + u\frac{4L^2}{q_1^2}\right)\frac{\| g_k \|}{\| d_{k-1} \|}.$$

由(1.1)式得，当 $k \geqslant 1$ 时，

$$d_k = - g_k - \beta_k^* \frac{d_{k-1}^T g_k}{\| g_k \|^2} g_k + \beta^* d_{k-1},$$

$$\| d_k \| \leqslant \| g_k \| + | \beta_k^* | \frac{| d_{k-1}^T g_k |}{\| g_k \|^2}\| g_k \| + | \beta_k^* | \| d_{k-1} \| = \| g_k \| + \left(\frac{2L}{q_1} + u\frac{4L^2}{q_1^2}\right)\| g_k \| + \left(\frac{2L}{q_1} + u\frac{4L^2}{q_1{}^2}\right)\| g_k \| = \left(1 + \frac{4L}{q_1} + u\frac{8L^2}{q_1^2}\right)\| g_k \|.$$

令 $q_2 = 1 + \frac{4L}{q_1} + u\frac{8L^2}{q_1^2}$，得到 $\| d_k \| \leqslant q_2 \| g_k \|$，证明完毕.

定理1 如果假设1，假设2和假设3都成立，考虑修正的PRP算法，则有

$$\liminf_{k\to\infty} \| g_k \| = 0. \tag{2.7}$$

证明 假设(2.7)式不成立，则存在常数 $\gamma > 0$，使得 $\| g_k \| > \gamma, \forall k > 0$，由(2.4)式和(2.6)式得到

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\| d_k \|^2} = \sum_{k=1}^{\infty} \frac{\| g_k \|^4}{\| d_k \|^2} \geqslant \sum_{k=1}^{\infty} \frac{\| g_k \|^2}{q_2^2} \geqslant \sum_{k=1}^{\infty} \frac{\gamma^2}{q_2^2} \to \infty.$$

该式与Zoutendijk条件相矛盾，所以假设不成立，证明完毕.

3 数值计算

用problem表示函数名，n 表示变量维数，nf 表示函数的迭代次数，$f(x)$ 表示函数的计算结果，$\| g(x) \|$ 表示梯度的计算结果，sut 表示迭代所用时间. 比较表1和表2，可以看出本文所研究的算法和PRP算法在上述函数中的计算效率是相当的，说明本算法对上述函数是有效的，而且具有良好的收敛性质和数值表现.

表1 算法1的数值结果

Table 1 Numerical result of algorithm 1

Problem	n	nf	$f(x)$	$\| g(x) \|$	sut
ARGLINA	200	5	0.20000E+03	0.26860E−13	0.03499
DQRTIC	5000	12	0.15125E+17	0.20061E+11	0.02100
FLETCBV3	5000	25	−0.15671E+15	0.14498E+04	0.07599
FLETCHBV	5000	22	−0.26415E+22	0.90844E+11	0.06799
NONCVXU2	5000	11	0.67434E+10	0.14125E+05	0.02799
NONCVXUN	5000	11	0.20112E+11	0.21383E+05	0.02899
PENALTY2	200	3	0.47116E+14	0.51779E+05	0.00200
QUARTC	5000	12	0.15125E+17	0.20061E+11	0.02100
TOINTGSS	5000	7	0.10002E+02	0.71897E−06	0.02799

表2 PRP算法的数值结果

Table 2 Numerical result of PRP algorithm

Problem	n	nf	$f(x)$	$\| g(x) \|$	sut
ARGLINA	200	5	0.20000E+03	0.26860E−13	0.03599
DQRTIC	3000	25	0.10000E+01	0.54338E−06	0.02300
FLETCBV3	5000	25	−0.15671E+15	0.14498E+04	0.07699
FLETCHBV	5000	22	−0.26415E+22	0.90844E+11	0.06699
NONCVXU2	5000	11	0.67434E+10	0.14125E+05	0.02999
NONCVXUN	5000	11	0.20112E+11	0.21383E+05	0.02899
PENALTY2	200	3	0.47116E+14	0.51779E+05	0.00200
QUARTC	5000	12	0.15125E+17	0.20061E+11	0.02100
TOINTGSS	5000	7	0.10002E+02	0.71897E−06	0.02799

参考文献:

[1] Hestenes M R, Stiefel E. Methods of conjugate gradients for

solving linear systems [J]. Journal of research, National Bureau of Standards: Section B, 1952, 49: 409-436.

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

[3] Polak E, Ribiere G. Note sur la convergence de directions conjugees [J]. Rev Prancaise Informat Recherche Operatinelle, 1969, 6: 35-43.

[4] Polyak B T. The conjugate gradient method in extreme problems [J]. USSR Computational Mathematics and Mathematical Physics, 1969, 9: 94-112.

[5] Hager W W, Zhang H. A new conjugate gradient method with guaranteeddescent and an efficient line search [J]. SIAM Journal on Optimization, 2005, 16: 170-192.

[6] Dai Y, Yuan Y. A nonlinear conjugate gradient with a strong global convergence properties [J]. SIAM Journal on Optimization, 2000, 10: 177-182.

[7] Yuan G L. Modified nonlinear conjugate gradiente methods with sufficient descent property for large-scale optimization problems [J]. Optimization Letters, 2009, 3: 11-21.

[8] Li Z, Zhou W J, Li D H. Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search [J]. Numerical Mathematics, 2006, 104: 561-572.

[9] Yuan G L, Lu X W, Wei Z X. A conjugate gradient method with descent direction forunconstrained optimization [J]. Journal of Computational and Applied Mathematics, 2009, 233: 519-530.

[10] Li G Y, Tang C M, Wei Z X. New conjugacy condition and related new conjugate gradient methods for unconstrained optimization [J]. Journal of Computational and Applied Mathematics, 2007, 202: 523-539.

[11] Yao S W, Wei Z X, Huang H. A note about WYL's conjugate gradient method and its applications [J]. Applied Mathematics and Computation, 2007, 191: 381-388.

[12] Dai Z F, Tian B S. Global convergence of some modified PRP nonlinear conjugate gradient methods [J]. Optimization Letters, 2011, 5: 615-630.

[13] Yuan G L, Wei Z X. Convergence analysis of a modified BFGS method on convex minimazations [J]. Computational Optimization and Applications, 2010, 47: 237-255.

(责任编辑：尹　闯)

(上接第98页 Continue from page 98)

[6] Liu G R, GU Y T. An introduction to meshfree methods and their programming [M]. Netherlands: Springer Verlag, 2005.

[7] Franke R. Scattered data interpolation: tests of some methods [J]. Math Comput, 1982, 38: 181-200.

[8] Qin Q H, Wang H, Kompis V. MFS with RBF for thin plate bending problems on elastic foundation [C]//Manolis G D, Polyzos D, eds. Recent Advances in Boundary Element Methods. New York: Spring Verlag, 2009, 38: 367-378.

[9] Chen W, Shen Z J, Yuan G W. General solutions and fundamental solutions of varied orders to the vibratioal thin, the Berger, and the Winkler plates [J]. Eng Anal Bound Elem, 2005, 29: 699-702.

(责任编辑：尹　闯)