第 8 卷第 4 期 2010 年 12 月
1672-6553/2010/08(4)/342-4

动力学与控制学报
JOURNAL OF DYNAMICS AND CONTROL

Vol. 8 No. 4
Dec. 2010

一种改进的共轭梯度法及全局收敛性*

朱铁锋[1,2]　刘雪英[1]

(1. 内蒙古工业大学数学系,呼和浩特市　010051)(2. 内蒙古师范大学青年政治学院计算机系,呼和浩特市　010051)

摘要　在 DY 共轭梯度法的基础上对解决无约束最优化问题提出一种改进的共轭梯度法. 该方法在标准 wolfe 线搜索下具有充分下降性,且算法全局收敛. 数值结果表明了该算法的有效性. 最后将算法用于 SO_2 氧化反应动力学模型的非线性参数估计,获得满意效果.

关键词　共轭梯度法,　充分下降性,　wolfe 线搜索,　全局收敛

引言

考虑如下无约束优化问题

$$\min f(x), x \in R^n \tag{1}$$

其中 $f: R^n \to R$ 为连续可微函数,在解决这类问题时的方法中,共轭梯度法因无需计算目标函数的二阶导数矩阵,所以是一类非常有效的算法[1-2]. 当目标函数连续可微时,其迭代格式为:

$$x_{k+1} = x_k + \alpha_k d_k, k = 1, 2\cdots, \tag{2}$$

$$d_k = \begin{cases} -g_k, & k = 1 \\ -g_k + \beta_k d_{k-1}, & k \geqslant 2 \end{cases} \tag{3}$$

其中 $g_k = \nabla f(x_k)$, d_k 为搜索方向,一般情况下,要求 d_k 满足 $g_k^T d_k < 0$,这样的方向 d_k 为函数 $f(x)$ 在 x_k 处的下降方向[3],更进一步要求 d_k 满足充分下降性

$$g_k^T d_k < -c \| g_k \|^2 \tag{4}$$

其中 $c > 0$ 为常数. α_k 为步长因子,可通过 Armijo 线搜索或 wolfe 线搜索求得. 常用的一种非精确线搜索是标准 wolfe 线搜索,其形式如下:

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \rho \alpha_k g_k^T d_k \tag{5}$$

$$g(x_k + \alpha_k d_k)^T d_k \geqslant \sigma g_k^T d_k \tag{6}$$

其中 $0 < \rho < \sigma < 1$. β_k 是一标量,著名的计算公式有:

$$\beta_k^{FR} = \frac{\| g_k \|^2}{\| g_{k-1} \|^2}([4], 1964)$$

$$\beta_k^{PR} = \frac{g_k^T (g_k - g_{k-1})}{\| g_{k-1} \|^2}([5], 1969)$$

$$\beta_k^{DY} = \frac{\| g_k \|^2}{d_{k-1}^T (g_k - g_{k-1})}([6] 1995)$$

文献[7]中,作者提出了一个新的参数 β_k 的取法,即:

$$\beta_k^{\mathrm{vprp}}(\mu) = \frac{\mu_1 (\| g_k \|^2 - |g_k^T g_{k-1}|)}{\mu_2 |g_k^T d_{k-1}| + \mu_3 \| g_{k-1} \|^2} \tag{7}$$

其中 $\mu_1 \in (0, +\infty)$, $\mu_2 \in [\mu_1 + \varepsilon_1, +\infty)$, $\mu_3 \in (0, +\infty)$ 和 ε_1 是任给的正常数. 该方法具有充分下降性和全局收敛性,同时取得了比较好的数值结果.

受到文献[7]的启发本文在 DY 方法的基础上,给出了一个新的参数 β_k 的取法,即:

$$\beta_k^{N}(\mu) = \frac{\mu_1 (g_k^T d_k - \| g_k \|)}{\mu_2 |g_k^T d_{k-1}| + |g_{k-1}^T d_{k-1}|} \tag{8}$$

其中 $\mu_1 \in (-1, 0)$, $\mu_2 \in [1, +\infty)$. 本文在标准 wolfe 条件下证明了此方法具有搜索方向充分下降性和全局收敛性.

1　新的共轭梯度算法及公式的充分下降性

算法 1.1

步骤 1:给出 $\varepsilon \geqslant 0$, $x_1 \in R^n$, $d_1 = -g_1$, $k := 1$ 若 $g_1 = 0$,则停.

步骤 2:用强 wolfe 线搜索(5)(6),求出步长 α_k.

步骤 3:$x_{k+1} = x_k + \alpha_k d_k$, $g_{k+1} = g(x_{k+1})$,若 $\| g_{k+1} \| \leqslant \varepsilon$,则停.

步骤 4:由公式(8)计算 β_{k+1},由(3)式求 d_{k+1}.

步骤 5:$k := k+1$. 转步 2.

为了证明方法的充分下降性和全局收敛性,对

2010-04-28 收到第 1 稿,2010-08-31 收到修改稿.
* 内蒙古工业大学重点科学研究项目(ZD200815)

目标函数$f(x)$作如下假设：

($H1$)$f(x)$在水平集 $L_0=\{x\in R^n:f(x)\leqslant f(x_1)\}$上有下界，其中 x_1 为初始点；

($H2$)$f(x)$在水平集 L_0 上连续可微，且 $g(x)$是 Lipschitz 连续的，即存在常数 $L>0$，使得

$$\|g(x)-g(y)\|\leqslant L\|x-y\|,\ \forall x,y\in L_0 \quad (9)$$

定理 1.1　设 x_1 为任意给定初始点，$f(x)$满足假设条件($H1$)和($H2$). 考虑方法(2)和(3)，其中 $\beta_k=\beta_k^N$，则对于任意 $k\geqslant 1$，都有

$$g_k^Td_k\leqslant-\frac{\mu_2}{\mu_2-\mu_1}\|g_k\|^2 \quad (10)$$

证明：　当 $k=1$ 时，$d_1=-g_1$，于是 $g_1^Td_1=-\|g_1\|^2$，(10)式显然成立.

当 $k\geqslant 2$ 时，$g_k^Td_k=g_k^T(-g_k+\beta_kd_{k-1})\leqslant$

$$-\|g_k\|^2+\frac{\mu_1(g_k^Td_k)}{\mu_2|g_k^Td_{k-1}|}|g_k^Td_{k-1}|.$$

于是

$$(1-\frac{\mu_1}{\mu_2})g_k^Td_k\leqslant-\|g_k\|^2$$

$$g_k^Td_k\leqslant-\frac{\mu_2}{\mu_2-\mu_1}\|g_k\|^2$$

2　全局收敛性

引理 2.1　设 x_1 为任意给定初始点，$f(x)$满足假设条件($H1$)和($H2$). 考虑方法(2)和(3)构成的迭代，其中 d_k 为下降方向，而步长因子 α_k 由 wolfe 线搜索(5)和(6)式求得，则对于任意 $k\in N$，Zoutendijk 条件成立，即

$$\sum_{k\geqslant 1}^{\infty}\frac{(g_k^Td_k)^2}{\|d_k\|^2}<+\infty \quad (11)$$

引理 2.1 的证明参考文献[8].

定理 2.1　假设目标函数$f(x)$满足假设条件($H1$)和($H2$)，设点列$\{x_k\}$由算法 2.1 产生，步长 α_k 由线搜索(5)和(6)式确定，则点列$\{x_k\}$或者 $g_k=0$ 对某个 k 成立，或者

$$\lim_{k\to+\infty}\inf\|g_k\|=0 \quad (12)$$

证明：若 $g_k=0$ 对某个 k 成立，则 x_k 为稳定点，定理得证. 否则，假设结论不真，则必存在常数 $\varepsilon>0$，对所有的 $k\in N$，使得

$$\|g_k\|\geqslant\varepsilon \quad (13)$$

$$(\beta_k^N)^2=\left(\frac{\mu_1(g_k^Td_k-\|g_k\|)}{\mu_2|g_k^Td_{k-1}|+|g_{k-1}^Td_{k-1}|}\right)^2\leqslant$$

$$\frac{\mu_1^2(g_k^Td_k)^2}{(g_{k-1}^Td_{k-1})^2}\leqslant\left(\frac{g_k^Td_k}{g_{k-1}^Td_{k-1}}\right)^2 \quad (14)$$

由(2)式可得 $d_k+g_k=\beta_kd_{d-1}$，两端取模平方，整理得

$$\|d_k\|^2=(\beta_k^N)^2\|d_{k-1}\|^2-2g_k^Td_k-\|g_k\|^2$$

利用(14)式，可得

$$\|d_k\|^2=(\beta_k^N)^2\|d_{k-1}\|^2-2g_k^Td_k-\|g_k\|^2\leqslant$$

$$\left(\frac{g_k^Td_k}{g_{k-1}^Td_{k-1}}\right)^2\|d_{k-1}\|^2-2g_k^Td_k-\|g_k\|^2$$

由于 $g_k^Td_k\neq 0$，$K\in N$，对上式两端除以$(g_k^Td_k)^2$，可得

$$\frac{\|d_k\|^2}{(g_k^Td_k)^2}\leqslant\frac{\|d_{k-1}\|^2}{(g_k^Td_k)^2}-\frac{2}{g_k^Td_k}-\frac{\|g_k\|^2}{(g_k^Td_k)^2}=$$

$$\frac{1}{(g_{k-1}^Td_{k-1})^2}\|d_{k-1}\|^2-(\frac{1}{\|g_k\|}+\frac{\|g_k\|}{g_k^Td_k})^2+$$

$$\frac{1}{\|g_k\|^2}\leqslant\frac{1}{(g_{k-1}^Td_{k-1})^2}\|d_{k-1}\|^2+\frac{1}{\|g_k\|^2} \quad (15)$$

由于 $d_1=-g_1$，所以$\frac{\|d_k\|^2}{(g_1^Td_1)^2}=\frac{1}{\|g_1\|^2}$. 于是，根据(15)式可得

$$\frac{\|d_k\|^2}{(g_k^Td_k)^2}\leqslant\sum_{i=1}^{k}\frac{1}{\|g_i\|^2} \quad (16)$$

根据(13)和(16)知：

$$\frac{\|d_k\|^2}{(g_k^Td_k)^2}\leqslant\frac{k}{\varepsilon^2}$$

即：

$$\frac{(g_k^Td_k)^2}{\|d_k\|^2}\geqslant\frac{\varepsilon^2}{k} \quad (17)$$

对(17)式两端取和得

$$\sum_{k\geqslant 1}\frac{(g_k^Td_k)^2}{\|d_k\|^2}\geqslant\sum_{k\geqslant 1}\frac{\varepsilon^2}{k}=+\infty \quad (18)$$

式(18)与引理 3.1 结论矛盾，定理得证.

3　算法测试

本节选用文献[9]中两个算例，并与 FR、PR、DY 共轭梯度法进行比较，IT 表示算法迭代次数，time 表示所用时间，f_{opt}表示目标函数的最优值.

测试函数(1)

$$f(x)=10(x_1^2-x_2)^2+(1-x_1)^2+9(x_4-x_3^2)^2+(1-x_3)^2+10.1((x_2-1)^2+(x_4-1)^2)+19.8(x_2-1)(x_4-1)$$

初始点为 $x_1=(-3,-1,-3,-1)$；选择的参数为 $\|\nabla f(x_k)\|\leqslant 10^{-3}$，$\rho=0.01$，$\sigma=0.1$，$\mu_1=-0.5$，$\mu_2=1.5$，最优解为(1,1,1,1)，最优值为0. 对不同的共轭梯度法进行比较，其结果见表1.

表1 不同算法对函数(1)的测试结果

Table 1 The test results of function (1) in different alogrithms

β_k	IT	$time/s$	f_{opt}
β_k^N	23	0.0589	$1.2492e-008$
β_k^{FR}	98	0.5499	$1.5759e-008$
β_k^{PR}	33	0.0990	$1.6692e-007$
β_k^{DY}	45	0.0690	$1.8735e-007$

测试函数(2)

$f(x)=\sum_{i=1}^{N/2}[(x_{2i}-x_{2i-1}^2)^2+(1-x_{2i-1})^2]$，$N=120$，$x_1=(-1,2,1,-1,2,1,\cdots,-1,2,1)$；选择的参数为 $\|\nabla f(x_k)\|\leqslant 10^{-2}$，$\rho=0.01$，$\sigma=0.1$，$\mu_1=-0.5$，$\mu_2=1.5$ 最优解为 $(1,1,\cdots,1)$，最优值为0. 对不同的共轭梯度法进行比较，其结果见表2.

表2 不同算法对函数(2)的测试结果

Table 2 The test results of function (2) in different alogrithms

β_k	IT	$time/s$	f_{opt}
β_k^N	9	23.2588	$5.1437e-006$
β_k^{FR}	19	56.3499	$1.4603e-005$
β_k^{PR}	11	32.1299	$4.1389e-004$
β_k^{DY}	14	43.3787	$6.2478e-005$

由以上数值实验表的比较，我们可以看出本算法的数值结果较好.

4 在 SO_2 氧化反应动力学模型参数估计中的应用

优化算法在非线性系统参数识别[10]，力学参数估计中有着广泛的应用[11,12].

文献[11]对 SO_2 催化氧化机理作了深入研究，推导动力学模型为：

$$r=\frac{K_1P_{O_2}^{1/2}}{K_2+K_3P_{SO_3}+\dfrac{P_{SO_3}}{P_{SO_2}}}\left[1-\frac{P_{SO_3}}{P_{SO_2}P_{O_2}^{1/2}K_P}\right] \tag{19}$$

式中：r 为反应速率；P_{O_2}、P_{SO_2}、P_{SO_3} 分别为 O_2、SO_2、SO_3 的分压；

$K_1=K_{01}\exp(-E_1/RT)$；

$K_2=K_{02}\exp(-E_2/RT)$；

$K_3=K_{03}\exp(-E_3/RT)$

$$K_p=\exp[2.3026(\frac{4812.3}{T}2,82541gT)+2.2884\times10^3T-7.012\times10^{-7}T^2+1.197\times10^{-10}T^3+2.23]$$

4.1 本文算法对模型参数的估计

本文根据文献[11]所提供的实验数据，采用本文算法对模型参数 K_{01}，K_{02}，K_{03}，E_1，E_2，E_3 进行估计. 参数估计原则是使式(20) EQS 达到最小，此为优化目标函数：

$$EQS=\sum_{i=1}^{M}\left[\frac{\hat{\gamma}_i-\gamma_i}{\gamma_i}\right] \tag{20}$$

式中：M—样本容量，γ_i，$\hat{\gamma}_i$—第 i 个样本的实验值与估算值. 算法的参数设置与算法测试中一致.

4.2 结果分析

表3的数据表明本文算法2.1优于文献[11]用Powell非线性优化方法和文献[12]用改进遗传算法(EGA)进行参数估计时所得的结果，比较如表3所示.

表3 不同优化模型参数估计值的比较

Table 3 Different optimal model parameter comparison

algorithm	K_{01}	K_{02}	K_{03}	E_1	E_2	E_3	EQS
Powell	0.152	8.18	0.221	62073	2384	18949	6.76
EGA	0.32	10.255	6.945	62610	1734.8	55694	3.926
Algorithm2.1	0.47	8.357	2.023	62252	2134.7	46853	3.327

参 考 文 献

1 Wang Ch Y, Du Sh Q, and Chen Y Y. Global convergence properties of three-term conjugate gradient method with new-type line search. *Journal of Systems Science and Complexity* ,2004, 17(3) ,412 ~ 420

2 Dai Y H, Han J Y, and Liu G H. Convergence properties of nonlinear conjugate gradient methods. *SIAM Journal of optimization* ,2000,10:,345 ~ 358

3 Nocedal. J. , Wright J. S. , Numerical optimization. New-York:Springer-Verlag,1999

4 Fletcher, R. , Reeves, C. Function minimization by conjugate gradients. *Computer Journal*,1964,7:149 ~ 154

5 Polak, B. T. The conjugate gradient method in extreme problems. *USSR Comput. Math . Math. Phys*,1969,9:94 ~ 112

6 Dai Y. H. Yuan Y X. A nonlinear conjugate gradient with a strong global convergence property. *SIAM Journal of Optimization*,2000,10:177 ~ 182

7 Yu G, Zhao Y, Wei Z. A modification of the PRP conjugate gradient method. Nanning: Department of Mathmatics and Information Science, Guangxi University, 2003

8 Zoutendijk, G., Nonlinear programming, computational methods. In: Integer and Nonlinear Programming (Adadie J, ed), 1970: 37 ~ 86

9 孙清滢,刘新海. 结合广义 Armijo 步长搜索的一类新的三项共轭梯度算法及其收敛特征. 计算数学, 2004, 26 (1): 25 ~ 36 (Sun Q Y, Liu X H. Global convergence results of a new three terms conjugate gradient method with generalized armijo step size rule. *Mathematic Numerica Sinica*, 2004, 26(1): 25 ~ 36 (in Chinese))

10 曾威,于德介. 一种基于小生境遗传算法的迟滞非线性系统参数识别方法. 动力学与控制学报, 2004, 2(1): 82 ~ 86 (Zeng W, Yu D J. A parameter identification method based on a niche genetic algorthm for hysteretic nonlinear system. *Journal of Dynamics and Control*, 2004, 2(1): 82 ~ 86 (in Chinese))

11 陈振兴,叶华,刘今. 铯-铷-钒系低温硫酸催化剂上氧化反应速率的机理解释. 化学反应工程与工艺. 2001, 17 (2): 119 ~ 123 (Chen Zh X, Ye H, Liu J. Study on the mechamism for the low temperature so2 oxidation with cs-rb-v sulfuric acid catalyst. *Chemical Reaction Engineering and Technology*, 2001, 17(2): 119 ~ 123 (in Chinese))

12 郑启富,陈德钊. 优进遗传算法及其在化工数据处理中的应用. 浙江大学学报(工学版), 2003, 37(3): 303 ~ 313 (Zheng Q F, Chen D ZH. Eugenic evolution genetic algorithm and its application for chemical engineering data processing. *Journal of Zhe Jiang University* (*Engineering Science*), 2003, 37(3): 303 ~ 313 (in Chinese))

GLOBAL CONVERGENCE OF A MODIFIED CONJUGATE GRADIENT METHOD *

Zhu Tiefeng[1,2] Liu Xueying[1]

(1. *Department of Mathematics*, *Inner Mongolia University of Technology*, *Hohhot* 010051, *China*)

(2. *Department of Computer Science*, *Youth College of Political Science of Inner Mongolia Normal University*, *Hohhot* 010051, *China*)

Abstract A modified conjugate method for unconstrained optimization problem was presented on the basis of DY conjugate gradient method. It is proved that the new formula is of full descent under the condition of the strong wolfe line search. At the same time, the new formula can support the global convergence. The numerical results show that the method is of great value. And the algorithm was applied to nonlinear parameter estimation of buring anteiso-dynamics model of sulfur dioxide acted on by caesium-rubidium-varadium low temperature sulfur acid catalyst, and oblaine satisfactory results were obtained.

Key words conjugate gradient method, sufficient descent property, wolfe line search, global convergence

Received 28 April 2010, revised 31 August 2010.

* The key Research Project in Inner Mongolia University of Technology (ZD200815)