广西科学 Guangxi Sciences 2010,17(4):321～323,336

3 种推广的 DY 共轭梯度法及其全局收敛性*

Global Convergence for 3 Extended DY Conjugate Gradient Methods

董晓亮,谢星星,侯志军,梅　燕

DONG Xiao-liang,XIE Xing-xing,HOU Zhi-jun,MEI Yan

(北方民族大学信息与计算科学学院,宁夏银川　750021)

(School of Information & Computation Science, Beifang University of Nationalities,Yinchuan, Ningxia,750021,China)

摘要:在标准 DY 共轭梯度方法的基础上提出以 β_k^{DY} 为界的 3 种杂交共轭梯度算法,在适当的条件下证明了这些方法是全局收敛的,并用数值实验检验其有效性. 初步的数值实验表明,3 种共轭梯度法比标准 DY 共轭梯度法更合适求解测试函数.

关键词:无约束优化　共轭梯度法　全局收敛性

中图法分类号:O224　　**文献标识码**:A　　**文章编号**:1005-9164(2010)04-0321-03

Abstract:Basing on DY method,three algorithms were proposed and they are the hybrid conjugate gradient methods in which β_k^{DY} is the upper bound of the parameter β_k for the unconstrained optimization. These given methods possess the global convergence under suitable conditions. In addition, numerical tests show these algorithms are effective.

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

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,其主要迭代公式为

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

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

其中 $g_k=\nabla f(x_k)$,d_k 是搜索方向,β_k 是参数,α_k 是步长因子. 它的选取是通过非精确搜索得到的,即参数满足 $0<\rho<\sigma<1$ 的 Wolfe 条件:

$$f(x_k)-f(x_k+\alpha_k d_k)\geqslant -\rho\alpha_k g_k^T d_k, \tag{3}$$

$$\sigma g_k^T d_k\leqslant g(x_k+\alpha_k d_k)^T d_k. \tag{4}$$

满足(3)式,再将(4)式替换为下式的线搜索被称为强 Wolfe 条件:

$$|g(x_k+\alpha_k d_k)^T d_k|\leqslant -\sigma g_k^T d_k. \tag{5}$$

著名的共轭梯度方法有 HS 方法、PRP 方法、FR 方法和 DY 方法[1]. 其中,DY 方法因其良好的理论性态与优越的数值性能,一度受到关注. 文献[2～6]提出了一些与 DY 方法相关的下降算法.

受上述方法的启发,我们在 DY 公式 $\beta_k^{DY}=\frac{\|g_k\|^2}{d_{k-1}^T(g_k-g_{k-1})}$ 的基础上构造出带参数 θ 的 β_k:

$$\beta_k^\theta=\frac{\|g_k\|^2-\theta g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}},\theta\geqslant 0. \tag{6}$$

其中 $y_{k-1}=g_k-g_{k-1}$(容易看出,在精确线搜索下 $g_k^T d_{k-1}=0$,故(6)式将还原为 β_k^{DY}),并结合 DY 方法的性质得到 3 种杂交共轭梯梯度算法.

1　共轭梯度算法

步骤 1　选定初始步:给出 $x_1\in R^n$,$d_1=-g_1$,$\theta=\theta_0$ 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,否则转步骤 3.

步骤 3　计算方向 $d_k=-g_k+\beta_k d_{k-1}$.

步骤 4　计算步长 α_k 满足 Wolfe 搜索(或强 Wolfe 搜索),令 $x_{k+1}=x_k+\alpha_k d_k$,置 $k=k+1$,转步骤 2.

收稿日期:2010-01-12

修回日期:2010-04-06

作者简介:董晓亮(1981-),男,硕士,讲师,主要从事最优化理论与方法研究。

* 宁夏高等学校科学研究项目(2009JY006),宁夏回族自治区精品课程《运筹学》建设项目,2010 年度北方民族大学信息与计算科学学院大学生创新性实验计划项目(xjcx201013)资助。

2 3种杂交算法及其收敛性分析

为分析收敛性,作以下必要的假设:

(H_1) $f(x)$ 在水平集 $L=\{x\in R^n \mid f(x)\leqslant f(x_0)\}$ 内有界;

(H_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

$$\| g(x)-g(y)\| \leqslant M\| x-y\|,\forall x,y\in D. \tag{7}$$

定理 1[1] (Zoutendijk 条件)假设(H_1),(H_2)成立,考虑一般方法(1)式和(2)式,α_k 满足(3)式和(4)式,对于任意 $k\in N$,有 $g_k^Td_k<0$,则 $\sum_{k=1}^{\infty}\frac{(g_k^Td_k)^2}{\| d_k\|^2}<+\infty$.

定理 2 假设(H_1),(H_2)成立,α_k 满足(3)式和(4)式,β_k 取(6)式,则对任意 $k\in N$,有充分下降条件 $g_k^Td_k<-c\| g_k\|^2$ 成立,其中 c 是某个常数.

证明 当 $n=1$ 时,$g_1d_1=-\| g_1\|^2<0$. 若 $n=k-1$ 时结论成立,则考虑 $n=k$ 时的情形:

$$g_k^Td_k=g_k^T(-g_k+\beta_kd_{k-1})=$$
$$\frac{\| \boldsymbol{g}_k\|^2d_{k-1}^Tg_{k-1}-\theta(g_k^Td_{k-1})^2}{d_{k-1}^Ty_{k-1}}. \tag{8}$$

另外,由(4)式知 $d_{k-1}^Ty_{k-1}\geqslant-(1-\sigma)g_{k-1}^Td_{k-1}>0$,结合(6)式和(8)式得 $g_k^Td_k<-(1-\sigma)^{-1}\| g_k\|^2$,结论成立.

注 1 若 $f(x)$ 是严格凸函数,则当 β_k 取(6)式时将产生不依赖于任何线搜索的下降方向.

以下,我们将结合式(6)中的 β_k^θ 对共轭梯度算法中步骤3的 β_k 采取如下3种杂交方法讨论.

杂交算法 1

$$\beta_k^{DY_1}=\begin{cases}\beta_k^\theta, 0\leqslant\theta g_k^Td_{k-1}\leqslant\min\{2,1/\sigma\}\| g_k\|^2,\\ \beta_k^{DY}, \text{else}.\end{cases} \tag{9}$$

定理 3 考虑一般方法(1)式和(2)式,α_k 满足(3)式和(4)式,若 β_k 取(9)式,则有 $|\beta_k^{DY_1}|\leqslant\beta_k^{DY}$.

证明 由(9)式知,当 $\beta_k=\beta_k^\theta$ 时,$0\leqslant\theta g_k^Td_{k-1}\leqslant\min\{2,1/\sigma\}\| g_k\|^2$. 当 $0<\sigma<1/2$ 时,$0\leqslant\theta g_k^Td_{k-1}\leqslant2\| g_k\|^2$. 由(6)式易知 $-\beta_k^{DY}<\beta_k=\beta_k^\theta<\beta_k^{DY}$;当 $1/2\leqslant\sigma<1$ 时,$0\leqslant\theta g_k^Td_{k-1}\leqslant\| g_k\|^2/\sigma$. 由(6)式易知 $(1-1/\sigma)\beta_k^{DY}<\beta_k=\beta_k^\theta<\beta_k^{DY}$.

其他情况下,$\beta_k=\beta_k^{DY}$. 综上所述,$|\beta_k^{DY_1}|\leqslant\beta_k^{DY}$. 结论成立.

杂交算法 2

借鉴文献[7]中对参数 β_k 施加惩罚项的策略,取 β_k 为

$$\beta_k^{DY_2}=\max\{0,\beta_k^\theta+\min(\mu_k,0)\}. \tag{10}$$

其中,$\mu_k=\frac{\| g_k\|^2-\theta g_k^Td_{k-1}}{d_{k-1}^Ty_{k-1}}\frac{\theta g_k^Td_{k-1}}{\| g_k\|^2}$.

定理 4 假设(H_1),(H_2)成立,α_k 满足(3)式和(4)式,β_k 取(10)式,则 $0\leqslant\beta_k^{DY_2}\leqslant\beta_k^{DY}$.

证明 在(4)式的基础上,对(10)式分以下3类情况讨论:

(Ⅰ)当 $\theta g_k^Td_{k-1}>\| g_k\|^2$ 时,易得 $\mu_k=\frac{\| g_k\|^2-\theta g_k^Td_{k-1}}{d_{k-1}^Ty_{k-1}}\frac{\theta g_k^Td_{k-1}}{\| g_k\|^2}<0$,从而 $\beta_k^{DY_2}=0$. 事实上,

$$\beta_k^\theta+\mu_k=\frac{\| g_k\|^2-\theta g_k^Td_{k-1}}{d_{k-1}^Ty_{k-1}}+$$
$$\frac{\| g_k\|^2-\theta g_k^Td_{k-1}}{d_{k-1}^Ty_{k-1}}\frac{\theta g_k^Td_{k-1}}{\| g_k\|^2}=$$
$$\frac{\| g_k\|^2-\theta g_k^Td_{k-1}}{d_{k-1}^Ty_{k-1}}(1+\frac{\theta g_k^Td_{k-1}}{\| g_k\|^2})<0.$$

(Ⅱ)当 $0<\theta g_k^Td_{k-1}\leqslant\| g_k\|^2$ 时,易得 $\mu_k\geqslant0$,从而 $\beta_k^{DY_2}=\beta_k^\theta<\beta_k^{DY}$.

(Ⅲ)当 $g_k^Td_{k-1}<0$ 时,易得 $\mu_k=\frac{\| g_k\|^2-\theta g_k^Td_{k-1}}{d_{k-1}^Ty_{k-1}}\frac{\theta g_k^Td_{k-1}}{\| g_k\|^2}<0$. 此时 $\beta_k^{DY2}=\beta_k^\theta+\mu_k$. 若 $\beta_k^\theta+\mu_k\leqslant0$,$\beta_k^{DY_2}=0$;而当 $\beta_k^\theta+\mu_k>0$ 时,则有

$$\beta_k^{DY_2}=\beta_k^\theta+\mu_k=\frac{\| g_k\|^2-\theta g_k^Td_{k-1}}{d_{k-1}^Ty_{k-1}}+$$
$$\frac{\| g_k\|^2-\theta g_k^Td_{k-1}}{d_{k-1}^Ty_{k-1}}\frac{\theta g_k^Td_{k-1}}{\| g_k\|^2}=\beta_k^{DY}-$$
$$\frac{|\theta g_k^Td_{k-1}|^2}{d_{k-1}^Ty_{k-1}\| g_k\|^2}<\beta_k^{DY}.$$

综合上述3种情况,结论成立.

定理 5 假设(H_1),(H_2)成立,考虑一般方法(1)式和(2)式,α_k 满足式(3)和式(4),若 $|\beta_k|\leqslant\beta_k^{DY}$,则由该方法推得 $\liminf_{k\to\infty}\| g_k\|=0$.

证明 假设结论不成立. 即存在 $\varepsilon>0$,当 k 充分大时,有 $\| g_k\|>\varepsilon$. 现将(2)式进行移项、两边取模平方,并在两边同时除以 $(g_k^Td_k)^2$,由定理3和定理4知 $|\beta_k^{DY_1}|\leqslant\beta_k^{DY}$,且根据文献[1]中 β_k^{DY} 的等价形式 $\beta_k=\frac{d_k^Tg_k}{d_{k-1}^Tg_{k-1}}$,有

$$\frac{\| d_k\|^2}{(g_k^Td_k)^2}=(\beta_k^2\| d_{k-1}\|^2-2g_k^Td_k-$$
$$\| g_k\|^2)/(g_k^Td_k)^2\leqslant\frac{\| d_{k-1}\|^2}{(g_{k-1}^Td_{k-1})^2}-(\frac{1}{\| g_k\|}+$$
$$\frac{\| g_k\|}{g_k^Td_k})^2+\frac{1}{\| g_k\|^2}=\frac{\| d_{k-1}\|^2}{(g_{k-1}^Td_{k-1})^2}+\frac{1}{\| g_k\|^2}. \tag{11}$$

对(11)式进行递推,并注意 $d_1=-g_1$,得到

$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leqslant \sum_{i=1}^{k} \|g_i\|^2 = \varepsilon^{-2} k$，则有 $\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = +\infty$. 这与定理 1 矛盾. 则假设不成立，定理 5 成立.

结合定理 3 和定理 4，杂交算法 1 和杂交算法 2 均全局收敛.

杂交算法 3

β_k 取为

$$\beta_k^{DY3} = \begin{cases} \beta_k^{\theta}, & 0 \leqslant g_k^T d_{k-1} \leqslant 2\|g_k\|^2/(1+\sigma), \\ \beta_k^{DY}, & \text{else}. \end{cases} \tag{12}$$

定理 6[3] 假设(H_1)，(H_2)成立，考虑一般方法(1)式和(2)式，α_k 满足(3)式和(4)式，若记 $\tau_k = \beta_k/\beta_k^{DY}$，当 $\tau_k \in [\frac{\sigma-1}{\sigma+1}, 1]$，则由该方法推得 $\liminf_{k\to\infty} \|g_k\| = 0$.

定理 7 假设(H_1)，(H_2)成立，α_k 满足(3)式和(4)式，β_k 取(12)式，则由该方法推得

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

证明 若 $0 \leqslant \theta g_k^T d_{k-1} \leqslant 2\|g_k\|^2/(1+\sigma)$，$\beta_k = \beta_k^{\theta}$，$\tau_k = 1 - \theta g_k^T d_{k-1}/\|g_k\|^2 \in [\frac{\sigma-1}{\sigma+1}, 1]$，其它情形时 $\beta_k = \beta_k^{DY}$，显然 $\tau_k \in [\frac{\sigma-1}{\sigma+1}, 1]$. 由定理 6 可知结论成立.

3 数值实验

用文献[8]的测试函数来检验本文算法的可行性和有效性. 为方便比较，侧重于考虑搜索步长时的数值稳定性.

实验结果如表 1 所示，其中各列意义为：Problem 为测试函数名称，Dim 为算例规模，DY 为标准的 DY 方法，DY_1 为情形 1 的方法((9)式)，θ_1 为相应参数，DY_2 为情形 2 的方法((10)式)，相应的参数 θ_2 统一设置为 3，DY_3 为情形 3 的方法((12)式)，θ_3 为相应参数. NI/NF/NG 表示迭代次数/函数调用次数/梯度调用次数. “Failed”表示迭代超过 1000 次.

考虑到数值稳定性，实验中步长策略均采取强 Wolfe 线搜索. 参数统一设置为 $\rho = 0.01$，$\sigma = 0.1$. 收敛精度为 $\varepsilon = 10^{-6}$，所有实验均不采取重开始策略. 从表 1 的 21 组测试函数可以看出，本文的 3 个算法在一定程度上比 DY 方法更适合求解大部分测试函数. 如何充分利用 DY 方法的收敛性和优越的数值表现，以及更加合理地选取参数，还有待进一步研究.

表 1 数值实验结果

Table 1 Results of numerical experiment

Problem	Dim (n)	DY (NI/NF/NG)	DY_1		DY_2 (NI/NF/NG)	DY_3	
			NI/NF/NG	θ_1		NI/NF/NG	θ_3
Rosenbrock	2	68/810/113	32/451/60	4	31/496/58	38/469/75	6
Freudenstein and Roth	2	16/38/26	16/38/26	4	12/32/24	16/38/26	2
Beale	2	57/260/88	18/88/30	4	23/195/36	21/142/34	6
Jenrich and Sampson	2	14/37/25	11/78/20	2	8/124/15	11/78/20	2
Helical valley	3	92/478/144	42/197/73	4	32/214/47	35/227/60	6
Bard	3	71/190/112	33/67/50	4	21/237/32	34/119/56	2
Gaussian	3	4/9/5	4/9/5	2	4/9/5	4/9/5	2
Gulf research and development	3	2/52/3	2/52/3	2	2/52/3	2/52/3	2
Box 3-dimensional	3	17/189/37	22/200/45	2	93/435/157	20/196/40	0.5
Singular powell	4	Failed	103/363/166	6	133/408/207	283/1034/464	6
Wood	4	754/9280/900	85/221/132	6	216/1523/339	108/755/180	6
Kowalik and Osbone	4	708/2714/1222	81/261/128	3	92/288/147	76/589/120	6
Biggs EXP6	6	255/873/408	117/621/189	1	98/440/162	111/511/173	4
Penalty 1	2	29/257/49	22/196/45	2	28/213/54	16/201/44	0.5
Penalty 2	4	371/2093/480	143/1480/278	3	115/1130/238	143/1480/278	3
Variably dimensional	50	10/52/36	10/52/36	2	10/52/36	10/52/36	2
Broyden tridiagonal	200	41/88/47	38/82/45	2	34/74/41	36/78/42	4
Linear-full rank	1000	1/3/3	1/3/3	2	1/3/3	1/3/3	2
Linear-rank 1	1000	1/3/3	1/3/3	2	1/3/3	1/3/3	2
Linear-rank 0	1000	1/3/3	1/3/3	2	1/3/3	1/3/3	2
Extended Rosenbrock	10000	113/ 876/181	45/ 450/79	2	39/ 225/67	45/ 450/79	2

(下转第 336 页 Continue on page 336)

[2] 殷庆瑞，祝炳和. 功能陶瓷和微观结构性能与制备技术[M]. 北京：冶金工业出版社，2005：130-192.

[3] 景晓宁，赵建华，何陵辉. 固相烧结后期晶粒和气孔拓扑生长演化的二维相场模拟[J]. 材料科学工程学报，2003，21：170-173.

[4] Chen L Q, Yang W. Computer simulation of the domain dynamics of a quenched system with a large number of nonconserved order parameters: The grain-growth kinetics[J]. Phys Rev B, 1994, 50: 15752-15756.

[5] Chen L Q, Fan D. Computer simulation model for coupled grain growth-application to Al_2O_3-ZrO_2 two-phase systems[J]. J Am Ceram Soc, 1996, 79 (5): 1163-1168.

[6] 高英俊，张海林，金星，等. 相场法研究硬质颗粒钉扎的两相晶粒长大过程[J]. 金属学报，2009，45(10)：1190-1198.

[7] Suwa Y, Saito Y, Onodera H. Phase-field simulation of recrystallization based on the unified subgrain theory [J]. Comput Mater Sci, 2008, 44: 286-295.

[8] Li Y L, Chen L Q. Temperature-strain phase diagram for $BaTiO_3$ thin films[J]. Appl Phys Lett, 2006, 88(7): 072905.

[9] Sleinbach I, Apel M. Phase-field simulation of rapid crystallization of silicon on substrate[J]. Mater Sci Eng A, 2007, 449-451, 96-98.

[10] Asp K, Agren J. Phase-field simulation of sintering and related phenomena[J]. Acta Mater, 2006, 54: 1241-1248.

[11] 景晓宁，倪勇，何陵辉，等. 陶瓷烧结过程空隙演化的二维相场模拟[J]. 无机材料学报，2002，17：1078-1082.

[12] 谢晖，孙军，杨刘晓，等. Mo 粉末烧结过程的相场模拟[J]. 铸造技术，2006，27：1078-1082.

[13] Lu G Q, Ishizaki L X (ed.). Grain boundary controlled properties of fine ceramics[M]. Elsevier Applied Science, London, 1992: 88-100.

[14] Kazaryan A, Wang Y, Patton Bruce R. Generalized phase field approach for computer simulation of sintering: incorporation of rigid-body motion[M]. Scripta Materialia, 1999, 41: 487-492.

[15] Fan Danan, Long QingChen. Diffusion-controlled grain growth in two-phase solids[J]. Acta Mater, 1997, 45: 3297-3310.

[16] Oono Y, Pori S. Computationally efficient modeling of ordering of quenched phases[J]. Phys Rev Lett, 1987, 58(8): 836-839.

[17] Lu G Q. Evolution of the pore structure of a ceramic powder compact during sintering[J]. Journal of Materials Pechnology, 1996, 59: 297-302.

[18] Cavus Falamaki, Mahdi Shafiee, Alireza Aghaie. Initial sintering stage pore growth mechanism applied to the manufacture of ceramic membrane supports[J]. Journal of the European Ceramic Society, 2004, 24: 2285-2292.

(责任编辑：邓大玉)

(上接第 323 页 Continue from page 323)

致谢

对所受的 3 项资助项目，以及项目组成员谢星星、侯志军和梅燕 3 位同学所做的工作表示感谢.

参考文献：

[1] 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.

[2] Dai Y H, Yuan Y X. A nonlinear conjugate gradient method with a strong global convergence property [J]. SIAM Journal of Optimization, 1999, 10(1): 177-182.

[3] Dai Y H, Yuan Y. An efficient hybrid conjugate gradient method for unconstrained optimization[J]. Annals of Operations Research, 2001, 103: 33-47.

[4] 戴志锋，陈兰平. 一种混合 HS-DY 共轭梯度法[J]. 计算数学，2005，27(4)：429-434.

[5] 董晓亮，李郴良，何郁波. 一类修正的 DY 共轭梯度法及其全局收敛[J]. 数值计算与计算机应用，2010，31(1)：1-7.

[6] 董晓亮，高岳林，何郁波. 一类混合的 FR-PC 共轭梯度法及其全局收敛性[J]. 河南师范大学学报，2010，38(2)：42-44.

[7] 莫降涛，顾能柱，韦增欣. 修正 PRP 共轭梯度法的全局收敛性及其数值结果[J]. 数值计算与计算机应用，2007，28(1)：56-62.

[8] Morè J J, Garbow B S, Hillstrome K E. Testing unconstrained optimization software[J]. ACM Trans Math Software, 1981, 7: 17-41.

(责任编辑：尹 闯)