研究简报

全局优化的非单调谱共轭梯度算法

马明娟[1]，梁　心[1]，黄庆道[2]
(1. 空军航空大学 基础部，长春 130022；2. 吉林大学 数学学院，长春 130012)

摘要：在非单调条件下给出一系列的谱共轭梯度算法，并根据不同的表达形式，给出了收敛性分析. 结果表明，该算法在迭代次数上明显优于其他算法.

关键词：全局优化；谱共轭梯度；非单调线搜索
中图分类号：O241.3　　**文献标志码**：A　　**文章编号**：1671-5489(2011)03-0475-03

Globally Optimized Spectral Conjugate Gradient Methods under Nonmonotone Conditions

MA Ming-juan[1], LIANG Xin[1], HUANG Qing-dao[2]
(1. *Department of Foundation, Aviation University of Air Force, Changchun* 130022, *China*;
2. *College of Mathematics, Jilin University, Changchun* 130012, *China*)

Abstract: A class of spectral conjugate gradient methods was proposed which consider a nonmonotone line search scheme. According to different formulae, we gave a convergence analysis of some numerical experiments. Numerical results show the new method is more efficient in the iterative number.

Key words: global optimization; spectral-conjugate gradient; nonmonotone line search

0　引　言

考虑如下无约束最优化问题：

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

其中：f: $R^n \to R$ 是连续可微函数；R^n 是 n 维欧几里德空间.

求解问题(1)的共轭梯度法一般采用如下迭代形式：

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \tag{2}$$

$$\boldsymbol{d}_k = \begin{cases} -\boldsymbol{g}_k, & k = 0, \\ -\boldsymbol{g}_k + \beta_k \boldsymbol{d}_{k-1}, & k > 0, \end{cases} \tag{3}$$

其中：$\boldsymbol{d}_k$ 是搜索方向；β_k 是一个标量；$\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$.

对于 β_k 的选择，文献[1-2]有许多经典形式. 文献[3-8]结合一般的 Wolfe 线搜索给出了算法的收敛性分析. 事实上，在非单调条件下的一些算法比单调条件下的算法数值表现更有效，尤其在牛顿和拟牛顿方法中[8-9]. 文献[10]在非单调条件下给出了一系列的共轭梯度算法. 本文在非单调条件下并结合谱的概念，给出了一系列谱共轭梯度(SCG)算法.

选取方向

$$\boldsymbol{d}_k = -\boldsymbol{g}_k + \beta_k \boldsymbol{d}_{k-1}. \tag{4}$$

收稿日期：2010-04-09.

作者简介：马明娟(1980—)，女，汉族，博士，讲师，从事优化理论和算法的研究，E-mail：mingjuanma@gmail.com. 通讯作者：黄庆道(1968—)，男，汉族，博士，教授，博士生导师，从事优化理论和生物数学的研究，E-mail：huangqd@jlu.edu.cn.

基金项目：国家自然科学基金(批准号：J0630104；40974047).

对于β_k的选取，以$\beta_k=\theta_k\beta_k^{HS}$为例给出收敛性结果. 对于其他形式的$\beta_k$，也可以得到相似结论. 这里

$$\theta_k = \boldsymbol{s}_{k-1}^{\mathrm{T}}\boldsymbol{s}_{k-1}/(\boldsymbol{s}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}), \tag{5}$$

其中：$\boldsymbol{s}_{k-1}=\boldsymbol{x}_k-\boldsymbol{x}_{k-1}$；$\boldsymbol{y}_{k-1}=\boldsymbol{g}_k-\boldsymbol{g}_{k-1}$.

假设：

(H_1) 水平集$L=\{\boldsymbol{x}\in R^n: f(\boldsymbol{x})\leqslant f(\boldsymbol{x}_1)\}$有界，其中$\boldsymbol{x}_1$为初始点；

(H_2) 水平集$L=\{\boldsymbol{x}\in R^n: f(\boldsymbol{x})\leqslant f(\boldsymbol{x}_1)\}$的一个$U$邻域内，$f$连续可微，导数$\boldsymbol{g}$满足Lipschitz条件，即存在常数$L>0$，使得$\|\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{g}(\boldsymbol{y})\|\leqslant L\|\boldsymbol{x}-\boldsymbol{y}\|$，$\forall \boldsymbol{x},\boldsymbol{y}\in U$.

称方向$\boldsymbol{d}_k$满足下降性条件是指其满足如下不等式：

$$\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k < 0, \qquad \forall k\geqslant 0. \tag{6}$$

1 SCG算法

下面给出SCG算法的迭代方法和表达形式.

初始化：选择$0<\lambda<\sigma<1$.

1) $\boldsymbol{d}_1=-\boldsymbol{g}_1$，$k\leftarrow 1$. 如果$\|\boldsymbol{g}_1\|=0$，终止算法，否则转2).

2) 计算步长α_k，$\alpha_{\mathrm{init}}=1$. 如果α满足：

$$f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)\leqslant \max_{0\leqslant j\leqslant m(k)} f(\boldsymbol{x}_{k-j})+\lambda\alpha\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k; \tag{7}$$

$$|\boldsymbol{g}(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)^{\mathrm{T}}\boldsymbol{d}_k|\leqslant -\sigma\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k, \qquad 0<\lambda<\sigma<1/2, \tag{8}$$

$$m(0)=0, \qquad m(k)\leqslant \min\{m(k-1)+1, M_0\}.$$

则定义$\alpha_k=\alpha$，终止线搜索转3)，否则，$\alpha_{\mathrm{new}}=\tau\alpha$，$\tau\in(0.1,0.9)$，$\alpha\leftarrow\alpha_{\mathrm{new}}$，回到这一步的开始.

3) $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k$，计算$\|\boldsymbol{g}_{k+1}\|=\|\nabla f(\boldsymbol{x}_{k+1})\|$. 如果$\|\boldsymbol{g}_{k+1}\|=0$，终止算法，并且输出$\boldsymbol{x}_{k+1}$，否则转4).

4) 计算θ_{k+1}和β_{k+1}，计算方向$\boldsymbol{d}=-\boldsymbol{g}_{k+1}+\beta_{k+1}\boldsymbol{d}_k$. 如果$\boldsymbol{d}^{\mathrm{T}}\boldsymbol{g}_{k+1}\leqslant\delta\|\boldsymbol{d}\|\,\|\boldsymbol{g}_{k+1}\|$，定义$\boldsymbol{d}_{k+1}=\boldsymbol{d}$；否则，定义$\boldsymbol{d}_{k+1}=-\theta_{k+1}\boldsymbol{g}_{k+1}$.

5) $k=k+1$，转2).

2 收敛性分析

引理1[10] 假设共轭梯度法满足非单调的线搜索条件和下降条件(6)，设$\eta_k=\lambda\alpha_k\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k$. 则序列$f(\boldsymbol{x}_{l(k)})$是单调递减的，其中：$l(k)=\max\{i\,|\,0\leqslant k-i\leqslant m(k); f(\boldsymbol{x}_i)=\max\limits_{0\leqslant j\leqslant m(k)} f(\boldsymbol{x}_{k-j})\}$.

引理2[10] 假设(H_1)成立，α_k满足线搜索条件(7),(8)及最速下降条件(6)，则$\lim\limits_{k\to\infty}\eta_{l(k+1)-1}=0$.

引理3[10] 假设条件(H_1),(H_2)成立，α_k由非单调线搜索条件(7),(8)得到，则存在常数c_1，使得$\|\boldsymbol{s}_k\|\geqslant\dfrac{c_1\langle -\boldsymbol{g}_{k-1},\boldsymbol{d}_{k-1}\rangle}{\|\boldsymbol{d}_{k-1}\|}$.

定理1 考虑具有形式(2)和(4)的共轭梯度法，这里α_k由非单调线搜索条件(7),(8)得到，假设条件(H_1),(H_2)成立，并且满足最速下降条件(6)，则

$$\liminf_{k\to\infty}\|\boldsymbol{g}_k\|=0. \tag{9}$$

证明：反证法. 假设$\lim\limits_{k\to\infty}\|\boldsymbol{g}_k\|>0$，则存在$\bar{\gamma}>0$，使得$0<\bar{\gamma}\leqslant\|\boldsymbol{g}(\boldsymbol{x})\|\leqslant\gamma$成立. 设$M=\max\{|c|,|d|\}$，则由$d\leqslant\theta_k\leqslant c$，可得

$$\theta_k=\frac{\boldsymbol{s}_{k-1}^{\mathrm{T}}\boldsymbol{s}_{k-1}}{\boldsymbol{s}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\leqslant c, \tag{10}$$

$$|\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}|\geqslant\frac{1}{M}\alpha_{k-1}^{-1}\|\boldsymbol{s}_{k-1}\|^2, \tag{11}$$

从而可得

$$\|\boldsymbol{g}_k^{\mathrm{T}}\theta_k\boldsymbol{y}_{k-1}\|\leqslant c\|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{y}_{k-1}\|\leqslant c\gamma\|\boldsymbol{y}_{k-1}\|\leqslant cL\gamma\|\boldsymbol{s}_{k-1}\|. \tag{12}$$

于是，由式(8),(10)~(12)，可得

$$\|\boldsymbol{d}_k\| \leqslant \|\boldsymbol{g}_k\| + |\beta_k|\,\|\boldsymbol{d}_{k-1}\| = \|\boldsymbol{g}_k\| + \left|\frac{\boldsymbol{g}_k^{\mathrm{T}}\theta_k\boldsymbol{y}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}}\right|\|\boldsymbol{d}_{k-1}\| \leqslant$$

$$\gamma + \frac{cML\gamma\,\|\boldsymbol{s}_{k-1}\|}{\alpha_{k-1}^{-1}\,\|\boldsymbol{s}_{k-1}\|^2}\|\boldsymbol{d}_{k-1}\| \leqslant \gamma(1 + cML). \tag{13}$$

又由引理 3 可知, 当 $k > k_0$ 时, 有

$$-\eta_{l(k+1)-1} = \frac{\lambda\,\|\boldsymbol{s}_{l(k+1)-1}\|\langle -\boldsymbol{g}_{l(k+1)-1}, \boldsymbol{d}_{l(k+1)-1}\rangle}{\|\boldsymbol{d}_{l(k+1)-1}\|} \geqslant \lambda c_1 \frac{\langle -\boldsymbol{g}_{l(k+1)-1}, \boldsymbol{d}_{l(k+1)-1}\rangle^2}{\|\boldsymbol{d}_{l(k+1)-1}\|^2} \geqslant \frac{\lambda c_1 \bar{c}^2\bar{\gamma}^4}{\|\boldsymbol{d}_{l(k+1)-1}\|^2};$$

由引理 2 可知,

$$\lim_{k\to\infty}\|1/\boldsymbol{d}_{l(k+1)-1}\|^2 = 0,$$

这与式(13)矛盾. 所以式(9)成立.

3　数值结果

下面用文献[11]中的测试函数检验本文的算法. 选取停止准则 $\|\boldsymbol{g}_k\| \leqslant 10^{-6}$. 表 1 列出了不同算法迭代次数的比较结果. 其中 n 表示变量的维数. 本文只对 $\beta_k = \theta_k\beta_k^{FR}$ 的形式进行简单说明.

表 1　不同算法迭代次数的比较结果

Table 1　Comparison results of different methods

n	SCG	GBB	CONMIN	PR^+
100	43	52	40	33
1 000	60	74	63	44
10 000	70	82	71	40

由表 1 可见, 本文算法的迭代次数明显优于其他算法.

参　考　文　献

[1] 袁亚湘, 孙文瑜. 最优化理论和方法 [M]. 北京: 科学出版社, 2007.

[2] Dai Y H, Yuan Y X. A Nonlinear Conjugate Gradient Methods with a Strong Global Convergence Property [J]. SIAM J Optim, 1999, 10: 177-182.

[3] YAN Hui, CHEN Lan-ping. Extend AS-GN Hybird Conjugate Gradient Method [J]. OR Transactions, 2010, 14(3): 122-128. (闫晖, 陈兰平. 推广 AS-GN 混合共轭梯度算法 [J]. 运筹学学报, 2010, 14(3): 122-128.)

[4] WANG Kai-rong, LIU Jin-kui, ZOU Li-min. Global Convergence of a Modified HS Conjugate Gradient Method [J]. Numerical Mathematics A Journal of Chinese University, 2010, 32(2): 150-156. (王开荣, 刘金魁, 邹黎敏. 一种修正的 HS 共轭梯度法和全局收敛性 [J]. 高等学校计算数学学报, 2010, 32(2): 150-156.)

[5] Armand P. Modification of the Wolfe Line Search Rules to Satisfy the Descent Condition in the PRP Conjugate Gradient Method [J]. JOTA, 2007, 132(2): 287-305.

[6] LI Guo-yi, TANG Chun-ming, WEI Zeng-xin. New Conjugacy Condition and Related New Conjugate Gradient Methods for Unconstrained Optimization [J]. Journal of Computational and Applied Mathematics, 2007, 202(2): 523-539.

[7] MA Ming-juan, DENG Jian, HUANG Qing-dao, et al. Spectral-Conjugate Gradient Mehtods with Inexact Line Search [J]. Journal of Jilin University: Science Edition, 2009, 47(2): 207-210. (马明娟, 邓键, 黄庆道, 等. 非精确条件下的谱共轭梯度算法 [J], 吉林大学学报: 理学版, 2009, 47(2): 207-210.)

[8] Toint P L. An Assessments of Nonmonotone Linesearh Technique for Unconstrained Optimization [J]. SIAM Journal on Scientific Computing, 1996, 17(3): 725-739.

[9] ZHANG Hong-chao, Hager W W. A Nonmonotone Line Search Technique and Its Application to Unconstrained Optimization [J]. SIAM Journal on Optimization, 2004, 14(4): 1043-1056.

[10] Liu G H, Jing L L. A Class of Nonmonotone Conjugate Gradient Methods for Unconstrained Optimization [J]. JOTA, 1999, 101(1): 127-140.

[11] Raydan M. The Barzilain and Borwein Gradient Methods for the Large Scale Unconstrained Minimization Problem [J]. SIAM Journal on Optimization, 1997, 7(1): 26-33.