DOI:10.13656/j.cnki.gxkx.2012.01.017

Armijo 线搜索修正 LS 共轭梯度法的收敛性*

Convergence Properties of the Modified LS Conjugate Gradient Method with Armijo Line Search

黄　海

HUANG Hai

(**广西民族师范学院数学与计算机科学系,广西崇左**　532200)

(Department of Mathematics and Computer Science,Guangxi Normal University for Nationalities,Chongzuo,Guangxi,532200,China)

摘要:基于修正 LS 共轭梯度法,给出合适的初始步长,使采用 Armijo 线搜索的迭代过程满足充分下降性.在较弱的条件下,证明算法具有全局收敛性和至少线性收敛速率.

关键词:共轭梯度法　Armijo 线搜索　全局收敛性　线性收敛速率

中图法分类号:O224　　**文献标识码**:A　　**文章编号**:1005-9164(2012)01-0007-03

Abstract:Appropriate condition for the initial stepsize is proposed based on the modified LS conjugate method,which make the sufficient descent property hold at each iteration with Armijo line search. Global convergence and linear convergence rate of the corresponding algorithm are proved under some mild conditions.

Key words:conjugate gradient method,Armijo line search,global convergence,linear convergence rate

非线性共轭梯度法是求解大规模优化问题 $\min\{f(x)\mid x\in R^n\}$ 的一类有效方法,其迭代格式为

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

$$d_k=\begin{cases}-g_k,k=1,\\-g_k+\beta_kd_{k-1},k>1,\end{cases} \tag{2}$$

其中 $f:R^n\rightarrow R$ 连续可微,$g_k=\nabla f(x_k)$,d_k 为迭代搜索方向,t_k 为迭代步长($t_k>0$),标量参数 β_k 有多种选择[1~6]. t_k 可以由各种线搜索产生,常用的强 Wolfe 线搜索条件[1] 为

$$f(x_k+t_kd_k)\leqslant f_k+\delta t_kg_k^Td_k, \tag{3}$$

$$|g(x_k+t_kd_k)^Td_k|\leqslant-\sigma g_k^Td_k, \tag{4}$$

其中 $\delta\in(0,0.5)$,$\sigma\in(\delta,1)$;Armijo 线搜索[7] 条件为

$$t_k=\max\{\rho^iw_k,i=0,1,2\cdots\}\text{ 满足(3)式}, \tag{5}$$

其中 $\rho\in(0,1)$,$L>0$,$w_k=\dfrac{-g_k^Td_k}{L\|d_k\|^2}$,$\|\cdot\|$ 表示 l_2 范数.

传统的共轭梯度法有 HS,FR,PRP,CD,LS,DY 等方法。其中,PRP 方法是数值结果最好的方法,但是在精确线搜索、弱 Wolfe 线搜索、强 Wolfe 线搜索条件下,对一般的目标函数均不能建立其收敛性[1].韦增欣等[2] 对 PRP 方法改进得到 WYL 方法,它在强 Wolfe 线搜索下对非凸目标函数全局收敛,而且数值结果优于 PRP 方法。将这一思想应用到对 HS 和 LS 方法的改进也很成功[3,4].最近,陆莎等[6] 研究 WYL 方法,获得更弱的收敛性条件,使 WYL 方法在 Armijo 线搜索下具有强收敛性.受文献[6]思想的启发,本文在 Armijo 线搜索时研究修正 LS 方法[3,4](MLS 方法)

$$\beta_{k+1}^{\mathrm{MLS}}=\frac{g_{k+1}^T\left(g_{k+1}-\dfrac{\|g_{k+1}\|}{\|g_k\|}g_k\right)}{-d_k^Tg_k}, \tag{6}$$

的初始步长和收敛性问题.

1　初始步长及算法

在共轭梯度法的收敛性分析中,充分下降条件

收稿日期:2011-01-07

修回日期:2011-05-23

作者简介:黄　海(1969-),男,副教授,主要从事最优化理论与方法的研究。

*广西壮族自治区教育厅科研项目(201012MS215)和广西民族师范学院科研项目(200909)资助。

$\exists c \in (0,1), \forall k \geqslant 1, g_k^T d_k \leqslant -c \| g_k \|^2$, (7) 有很重要的作用. 对(1),(2) 和(6) 式的迭代法,文献[3,4]通过限制强 Wolfe 条件(4) 式中的参数 $\sigma \in (0, 0.5)$ 而使(7) 式满足,我们借鉴文献[6]的思想,忽略线搜索条件,仅考虑(1) 式中怎样的 t_k 才能使(7) 式满足,从而得到第 k 次迭代的初始步长为

$$\alpha_k = \frac{c(3-c)}{2L} \frac{\| g_k \|^2}{\| d_k \|^2}, \tag{8}$$

其中 $c \in (0,1)$,L 为 $f(x)$ 的梯度函数 $g(x)$ 的 Lipschitz 常数. 以(8) 式为初始步长采用 Armijo 线搜索的 MLS 算法定义如下.

算法 1

步骤 0　**给定初值** $x_1 \in R^n$,$\delta \in (0, 0.5)$,$c, \rho \in (0,1)$. **令** $d_1 = -g_1$,$\alpha_1 = (2L)^{-1} c(3-c)$,$k=1$.

步骤 1　**若** $\| g_k \| = 0$,**停止.**

步骤 2　Armijo **线搜索**

步骤 2.1　$t_k = 1$;

步骤 2.2　**若**

$$f(x_k + t_k \alpha_k d_k) - f_k \leqslant \delta t_k \alpha_k g_k^T d_k, \tag{9}$$

则令 $x_{k+1} = x_k + t_k \alpha_k d_k$,转步骤 3;否则,令 $t_k = \rho t_k$,转步骤 2.2.

步骤 3　**若** $\| g_{k+1} \| = 0$,**停止;否则,由**(6) **式和**(2) **式计算** d_{k+1},**由**(8) **式计算初始步长** α_{k+1}.

步骤 4　**令** $k = k+1$,**转步骤** 2.

2　算法的收敛性

假设(H1):$f(x)$ **在水平集** $\Omega = \{x \in R^n \mid f(x) \leqslant f(x_1)\}$ **上有下界.**

(H2):$f(x)$ **在** Ω **上连续可微,其梯度函数** $g(x)$ **在** Ω **上** Lipschitz **连续,即** $\exists L > 0$ **使**

$$\| g(y) - g(x) \| \leqslant L \| y - x \|, \forall x, y \in \Omega. \tag{10}$$

定理 1　**设**(H1) **和**(H2) **成立**,$\{g_k, d_k\}$ **为算法** 1 **生成的无穷序列,则** $\forall k \geqslant 1$,**有**

(i) $g_k^T d_k \leqslant -c \| g_k \|^2$;

(ii) $\| d_k \| \leqslant (4-c) \| g_k \|$.

证明 (i)　**显然** $g_1^T d_1 = -\| g_1 \|^2 \leqslant -c \| g_1 \|^2$,**以下假设** $g_k^T d_k \leqslant -c \| g_k \|^2$.

由(6) **式和** Cauchy-Schwartz **不等式,可得**

$$0 \leqslant \beta_{k+1}^{\mathrm{MLS}} \leqslant \frac{2 \| g_{k+1} \|^2}{c \| g_k \|^2}, \tag{11}$$

由(2) **式可得**

$$g_{k+1}^T d_{k+1} = -\| g_{k+1} \|^2 + \beta_{k+1}^{\mathrm{MLS}} g_{k+1}^T d_k, \tag{12}$$

若 $g_{k+1}^T d_k \leqslant 0$,**由**(11) **式和**(12) **式,可以知**

$$g_{k+1}^T d_{k+1} \leqslant -\| g_{k+1} \|^2 \leqslant -c \| g_{k+1} \|^2.$$

若 $g_{k+1}^T d_k > 0$,**由** Cauchy-Schwartz **不等式,**(10) **式和**(8) **式,可得**

$$g_{k+1}^T d_k = (g_{k+1} - g_k)^T d_k + g_k^T d_k \leqslant L t_k \alpha_k \| d_k \|^2 - c \| g_k \|^2 \leqslant 2^{-1} c(1-c) \| g_k \|^2,$$

再由(12) **式,**(11) **式及上式,可得**

$$g_{k+1}^T d_{k+1} \leqslant -c \| g_{k+1} \|^2.$$

综上,根据数学归纳法可知,$\forall k \geqslant 1$,(i) **成立.**

(ii) **显然** $\| d_1 \| = \| g_1 \| \leqslant (4-c) \| g_1 \|$.

$$\left\| g_{k+1} - \frac{\| g_{k+1} \|}{\| g_k \|} g_k \right\| = \frac{1}{\| g_k \|} \| g_k (\| g_k \| - \| g_{k+1} \|) + \| g_k \| (g_{k+1} - g_k) \| = | (\| g_k \| - \| g_{k+1} \|) | + \| g_{k+1} - g_k \| \leqslant 2 \| g_{k+1} - g_k \| \leqslant 2 L t_k \alpha_k \| d_k \|. \tag{13}$$

由(6) **式,利用** Cauchy-Schwartz **不等式,**(13) **式,**(8) **式及**(i),**可得**

$$0 \leqslant \beta_{k+1}^{\mathrm{MLS}} \leqslant (3-c) \| g_{k+1} \| (\| d_k \|)^{-1}. \tag{14}$$

由(2) **式和**(14) **式知,**

$$\| d_{k+1} \| - \| g_{k+1} \| \leqslant \| d_{k+1} - g_{k+1} \| = \beta_{k+1}^{\mathrm{MLS}} \| d_k \| \leqslant (3-c) \| g_{k+1} \|,$$

从而得到 $\| d_{k+1} \| \leqslant (4-c) \| g_{k+1} \|$.

综上可知,$\forall k \geqslant 1$,(ii) **成立. 定理得证.**

引理 1　**设**(H1) **和**(H2) **成立**,$\{t_k\}$ **为算法** 1 **生成的无穷序列,则**

$\exists \eta \in (0,1]$,**使** $\forall k \geqslant 1, t_k \in [\eta, 1]$.

证明　**设** $K_1 = \{k \mid t_k = 1\}$,$K_2 = \{k \mid t_k < 1\}$. **由算法** 1 **的步骤** 2 **知**,$\forall k \geqslant 1$,**若** $k \notin K_1$,**则** $k \in K_2$.

$\forall k \in K_2$,**由算法** 1 **的步骤** 2.2 **知**,$\rho^{-1} t_k$ **不满足**(9) **式,即**

$$f(x_k + \rho^{-1} t_k \alpha_k d_k) - f_k > \delta \rho^{-1} t_k \alpha_k g_k^T d_k. \tag{15}$$

由微分中值定理,Cauchy-Schwartz **不等式及**(10) **式知**,$\exists \theta_k \in (0,1)$,**使**

$$f(x_k + \rho^{-1} t_k \alpha_k d_k) - f_k = \rho^{-1} t_k \alpha_k [g(x_k + \theta_k \rho^{-1} t_k \alpha_k d_k) - g_k]^T d_k + \rho^{-1} t_k \alpha_k g_k^T d_k \leqslant \rho^{-2} t_k^2 \alpha_k^2 \theta_k L \| d_k \|^2 + \rho^{-1} t_k \alpha_k g_k^T d_k \leqslant 2^{-1} c(3-c) \rho^{-2} t_k^2 \alpha_k \| g_k \|^2 + \rho^{-1} t_k \alpha_k g_k^T d_k. \tag{16}$$

由(16) **式,**(15) **式及定理** 1,**可得**

$$2^{-1} c(3-c) \rho^{-1} t_k \| g_k \|^2 \geqslant -(1-\delta) g_k^T d_k \geqslant (1-\delta) c \| g_k \|^2,$$

从而得到 $t_k \geqslant 2\rho(1-\delta)(3-c)^{-1}$.

综上可知,$\forall k \geqslant 1$,$t_k \geqslant \min \{1, 2\rho(1-\delta)(3-c)-1\} \triangleq \eta$. **引理得证.**

定理 2　**设**(H1) **和**(H2) **成立**,$\{g_k\}$ **为算法** 1 **生成的无穷序列,则** $\lim\limits_{k \to \infty} \| g_k \| = 0$.

证明　**由**(9) **式,定理** 1 **及**(H1) **知**,$\{f_k\}$ **为严格单调递减且有界序列,**

$$\lim_{k\to\infty} f_k = f^* < +\infty, \forall k \geqslant 1, x_k \in \Omega.$$

由(9)式,(8)式,引理1和定理1,可得

$$f_k - f_{k+1} \geqslant -\delta t_k \alpha_k g_k^T d_k \geqslant \delta\eta \frac{c(3-c)}{2L} \frac{\| g_k \|^4}{\| d_k \|^2} \geqslant$$

$$\delta\eta \frac{c(3-c)}{2L(4-c)} \| g_k \|^2 \triangleq \sigma \| g_k \|^2, \tag{17}$$

上式两边分别求和,可得

$$+\infty > f_1 - f^* = \sum_{k=1}^{\infty} (f_k - f_{k+1}) \geqslant \sigma \sum_{k=1}^{\infty} \| g_k \|^2,$$

从而得到 $\lim_{k\to\infty} \| g_k \| = 0$,定理得证.

3 算法的收敛速率

假设(H3): $f(x)$ 在 R^n 上二次连续可微且一致凸.

引理 2[8] 若(H3)成立,则(H1)和(H2)成立,$f(x)$ 在 R^n 上有唯一的极小点 x^*,且 $\exists M > m > 0$,使 $\forall x \in R^n$,有

$$m \| x - x^* \| \leqslant \| g(x) \| \leqslant M \| x - x^* \|, \tag{18}$$

$$2^{-1} m \| x - x^* \|^2 \leqslant f(x) - f(x^*) \leqslant 2^{-1} M \| x - x^* \|^2. \tag{19}$$

定理 3 若(H3)成立,$\{x_k\}$ 为算法1生成的无穷序列,则 $\{x_k\}$ 至少 R-线性收敛于 x^*.

证明 由引理2及定理2知,$f(x)$ 有唯一极小点 x^*,$\lim_{k\to\infty} x_k = x^*$,$\lim_{k\to\infty} f_k = f(x^*) \triangleq f^*$,$f_k - f^* \geqslant 0$. 令 $\bar{m} = \min\{m, L\}$,由(17),(18)和(19)式,可得

$$f_k - f_{k+1} \geqslant \sigma \| g_k \|^2 \geqslant \sigma m^2 \| x_k - x^* \|^2 \geqslant 2\sigma \bar{m}^2 M^{-1} (f_k - f^*) \triangleq \tau (f_k - f^*). \tag{20}$$

因 $\eta, c \in (0,1)$,$\delta \in (0, 0.5)$,由(17)式可得 $\sigma < (4L)^{-1}$,又由常数 $\bar{m}, L, m, M$ 的关系,可知

$$0 < \tau = 2\sigma \bar{m}^2 M^{-1} \leqslant 2\sigma L \leqslant 2^{-1},$$

利用(20)式,递推可得

$$f_k - f^* = (f_k - f_{k-1}) + (f_{k-1} - f^*) \leqslant (1-\tau)(f_{k-1} - f^*) \leqslant (1-\tau)^{k-1}(f_1 - f^*),$$

由(19)式及上式,可得

$$\| x_k - x^* \|^2 \leqslant 2m^{-1}(f_k - f^*) \leqslant 2m^{-1}(f_1 - f^*)(1-\tau)^{k-1},$$

从而可得 $\lim_{k\to\infty} \| x_k - x^* \|^{\frac{1}{k}} \leqslant \sqrt{1-\tau} < 1$,说明 $\{x_k\}$ 至少 R-线性收敛于 x^*. 定理得证.

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.

[2] Wei Z X, Yao S W, Liu L Y. The convergence properties of some new conjugate gradient methods[J]. Appl Math Comput, 2006, 183: 1341-1350.

[3] Yao S W, Wei Z X, Huang H. A note about WYL's conjugate gradient method and its applications[J]. Appl Math Comput, 2007, 191: 381-388.

[4] 黄海,林穗华,姚胜伟. 一个基于LS公式修正的新共轭梯度算法[J]. 广西科学,2007,14(3):224-226.

[5] 董晓亮,谢星星,侯志军,等. 3种推广的DY共轭梯度法及其全局收敛性[J]. 广西科学,2010,17(4):321-323.

[6] Lu S, Wei Z X, Mo L L. Some global convergence properties of the Wei-Yao-Liu conjugate gradient method with inexact line search[J]. Appl Math Comput, 2011, 217: 7132-7137.

[7] Shi Z J, Shen J. Convergence of descent method without line search[J]. Appl Math Comput, 2005, 167: 94-107.

[8] Cohen A I. Stepsize analysis for descent methods[J]. J Optim Theory Appl, 1981, 33(2): 187-205.

(责任编辑:陈小玲)

美国研制成功高效存储氢的纳米复合材料

上世纪70年代,人们开始将氢气看成化石燃料的替代品并对其寄予厚望,因为氢气燃烧后得到的副产品只有水,而其他碳氢化合物燃料燃烧后会喷射出温室气体和有害污染物。另外,同汽油相比,氢气的质量更轻,能量密度更大而且来源丰富。但氢气要想作为燃料替代汽油,就必须解决两大难题:如何安全且密集地存储,以及如何更容易获得。科学家一直尝试解决这两个问题。他们试着将氢气"锁"在固体中;试着在更小的空间内存储更多氢气,同时让氢气的反应性很低(要让氢气这种易挥发的物质保持稳定,低反应性非常重要)。然而,大多数固体只能吸收少量氢气,同时,还需要对整个系统进行极度地加热或冷却来提升其能效。最近,美国科学家设计出了一种新的纳米储氢复合材料,其由金属镁纳米离子散落在一个聚甲基丙烯酸甲酯(同树脂玻璃有关的聚合物)基质组成。新材料在常温下就能快速地吸收和释放氢气,在吸收和释放氢气的循环中,金属镁也不会氧化。这是氢气储存和氢燃料电池等领域取得的又一个重大突破。

(据科学网)