第 32 卷第 5 期　　重庆工商大学学报(自然科学版)　　2015 年 5 月
Vol. 32　NO.5　　J Chongqing Technol Business Univ. (Nat Sci Ed)　　May.2015

doi:10.16055/j.issn.1672-058X.2015.0005.004

一个修正的三项 LS 共轭梯度法

焦佳佳

(重庆师范大学 数学学院,重庆 401331)

摘　要:基于已有的共轭梯度法的思想,提出了一个三项 LS 共轭梯度方法,该方法能保证搜索方向在不需要任何线搜索下具有充分下降性,并在适当条件下获得此方法对一般函数的全局收敛性.

关键词:共轭梯度法;充分下降性;全局收敛性

中图分类号:O224　　**文献标识码**:A　　**文章编号**:1672-058X(2015)05-0013-04

1　基础知识

考虑无约束优化问题:

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

其中 $f:\mathbf{R}^n \to \mathbf{R}$ 是连续可微函数.共轭梯度法是求解大规模无约束优化问题(1)的一种有效方法,其迭代格式如下:

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

$$d_k = \begin{cases} -g_k, k = 0 \\ -g_k + \beta_k d_{k-1}, k \geqslant 1 \end{cases} \tag{3}$$

其中 $g_k = \nabla f(x_k)$,d_k 是搜索方向,α_k 是用某线搜索得到的步长,β_k 是参数.不同的参数表达式 β_k 对应着不同的共轭梯度法,著名的有 β_k^{FR},β_k^{PRP},β_k^{HS} 等.常用的线搜索有 Wolfe 线搜索、强 Wolfe 线搜索、Armijo-type 线搜索等,其中 Wolfe 线搜索要求 α_k 满足

$$f(x_k + \alpha_k d_k) - f(x_k) \leqslant \delta \alpha_k g_k^{\mathrm{T}} d_k \tag{4}$$

$$g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k > \sigma g_k^{\mathrm{T}} d_k \tag{5}$$

Armijo-type 线搜索要求 α_k 满足

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) - \delta \alpha_k^2 \| d_k \|^2 \tag{6}$$

与其他共轭梯度法相比,PRP 方法的数值表现是目前认为最好的方法之一,但该方法的收敛性不理想.研究发现,PRP 方法收敛性不好的另一个原因在于它不具有充分下降性.充分下降性是指对所有的 k,式(7)成立:

$$g_k^{\mathrm{T}} d_k \leqslant -c \| g_k \|^2 \tag{7}$$

其中 c 为常数,此性质在收敛性分析中起着非常重要的作用.因此,许多学者都希望找到数值表现可与 PRP 相媲美同时性质又比其好的方法.

Yu,Guan 和 Li 在文献[4]中提出了具有充分下降性的修正 PRP 方法,其公式为

收稿日期:2014-08-10;**修回日期**:2014-10-08.

作者简介:焦佳佳(1988-),女,陕西铜川人,硕士研究生,从事最优化方法及其应用研究.

$$\beta_k^{\mathrm{MPRP}}=\beta_k^{\mathrm{PRP}}-\frac{\mu\|y_{k-1}\|^2}{\|g_{k-1}\|^4}g_k^{\mathrm{T}}d_{k-1} \tag{8}$$

其中 $\mu>\frac{1}{4}$ 是常数.该方法能保证 $\beta_k^{\mathrm{MPRP}}\geqslant 0$,且不依赖于任何线搜索具有充分下降性,所以其数值表现优于 PRP 方法.

Zhang 等在文献[5]中给出一个三项的 PRP 共轭梯度法,其方向定义为

$$d_k=\begin{cases}-g_k,k=0\\-g_k+\beta_k^{\mathrm{PRP}}d_{k-1}-\vartheta_{k-1}y_{k-1},k\geqslant 1\end{cases} \tag{9}$$

其中 $\vartheta_{k-1}=\frac{g_k^{\mathrm{T}}d_{k-1}}{\|g_{k-1}\|^2}$,$\beta_k^{\mathrm{PRP}}=\frac{g_k^{\mathrm{T}}y_{k-1}}{\|g_{k-1}\|^2}$,容易得到 $g_k^{\mathrm{T}}d_k=-\|g_k\|^2$.

与三项 PRP 方法类似,具有充分下降性的三项 LS 公式定义为

$$d_k=\begin{cases}-g_k,k=0\\-g_k+\beta_k^{\mathrm{LS}}d_{k-1}-\frac{g_k^{\mathrm{T}}d_{k-1}}{g_{k-1}^{\mathrm{T}}d_{k-1}}y_{k-1},k\geqslant 1\end{cases} \tag{10}$$

此方法仅仅利用了梯度值信息,忽略了函数值信息.文献[6]中给出了新的拟牛顿方程:

$$B_ks_{k-1}=y_{k-1}^*$$

其中 $y_{k-1}^*=y_{k-1}+t\|g_{k-1}\|^r s_{k-1}(t>0)$.

此拟牛顿方程同时含有函数值信息和梯度值信息,且比原拟牛顿方法具有更好的收敛性和数值表现.受此启发,此处给出一个三项 LS 共轭梯度方法,其方向定义为

$$d_k=\begin{cases}-g_k,k=0\\-g_k+\beta_k^{\mathrm{MLS}}d_{k-1}-\frac{g_k^{\mathrm{T}}d_{k-1}}{g_{k-1}^{\mathrm{T}}d_{k-1}}y_{k-1}^*,k\geqslant 1\end{cases} \tag{11}$$

其中 $\beta_k^{\mathrm{MLS}}=\frac{g_k^{\mathrm{T}}y_{k-1}^*}{-g_{k-1}^{\mathrm{T}}d_{k-1}}$.可以看出新公式同时含有梯度值信息和函数值信息.

2 算法及其全局收敛性分析

为表述需要,称修正的三项 LS 共轭梯度算法为 MMLS 算法,其步骤如下:

步 0:给出初值 $x_1\in\mathbf{R}^n$,$\varepsilon\geqslant 0$,令 $d_1=-g_1$,$k=1$;

步 1:若 $\|g_1\|\leqslant\varepsilon$,则停止,否则转步 2;

步 2:步长 $\alpha_k>0$,由式(6)获得;

步 3:令 $x_{k+1}=x_k+\alpha_kd_k$,$g_{k+1}=g(x_{k+1})$,若 $\|g_{k+1}\|\leqslant\varepsilon$,则停止,否则转步 4;

步 4:计算 β_k,利用式(11)计算 d_{k+1};

步 5:令 $k=k+1$,转步 2.

为了讨论新方法的全局收敛性,假定目标函数满足如下基本假设:

假设(A) 水平集 $\Omega=\{x\in\mathbf{R}^n|f(x)\leqslant f(x_1)\}$ 有界,即存在常数 $a>0$,使得

$$|x|\leqslant a,\forall x\in\Omega \tag{12}$$

假设(B) $f(x)$ 在水平集 Ω 的一个领域 N 内连续可微,且其梯度 $g(x)$ 满足 Lipschitz 条件,即存在常数 $L>0$,使得

$$\|g(x)-g(y)\|\leqslant L\|x-y\|,\forall x,y\in N \tag{13}$$

根据上述假设,存在一个正的常数 Γ,对任意的 $x\in\Omega$,都有

$$\|g(x)\|\leqslant \Gamma \tag{14}$$

成立.

引理 1　对 $k\geqslant 0$,三项 LS 公式的搜索方向满足

$$g_k^{\mathrm{T}}d_k=-\|g_k\|^2 \tag{15}$$

证明　如果 $k=0$,则 $g_0^{\mathrm{T}}d_0=-\|g_0\|^2$,那么式(15)成立.当 $k\geqslant 1$ 时,利用新方向的定义可得

$$g_{k+1}^{\mathrm{T}}d_{k+1}=-\|g_{k+1}\|^2+\frac{g_{k+1}^{\mathrm{T}}y_k^*}{-g_k^{\mathrm{T}}d_k}d_k^{\mathrm{T}}g_{k+1}-\frac{g_{k+1}^{\mathrm{T}}d_k}{-g_k^{\mathrm{T}}d_k}g_{k+1}^{\mathrm{T}}y_k^*=-\|g_{k+1}\|^2$$

证毕.

引理 2　如果存在一个常数 $\bar{\gamma}>0$,使得

$$\|g_k\|\geqslant \bar{\gamma},\ \forall k \tag{16}$$

则存在一个常数 $M>0$,使得

$$\|d_k\|\leqslant M,\ \forall k \tag{17}$$

证明　考虑 y_k^* 的定义,由式(13)和(14)知

$$\|y_{k-1}^*\|\leqslant \|y_{k-1}\|+t\|g_{k-1}\|^r\|s_{k-1}\|\leqslant L\|s_{k-1}\|+t\Gamma^r\|s_{k-1}\|=(L+t\Gamma^r)\|s_{k-1}\| \tag{18}$$

根据 d_k 的定义,由式(13)(14)(16)和(18)可得

$$\begin{aligned}\|d_k\|&=\left\|-g_k+\frac{g_k^{\mathrm{T}}y_{k-1}^*}{-g_{k-1}^{\mathrm{T}}d_{k-1}}d_{k-1}+\frac{g_k^{\mathrm{T}}d_{k-1}}{g_{k-1}^{\mathrm{T}}d_{k-1}}y_{k-1}^*\right\|\leqslant\\&\|-g_k\|+2\frac{\|g_k\|\|y_{k-1}^*\|\|d_{k-1}\|}{-g_{k-1}^{\mathrm{T}}d_{k-1}}=\\&\|-g_k\|+2\frac{\|g_k\|\|y_{k-1}^*\|\|d_{k-1}\|}{\|g_{k-1}\|^2}\leqslant\\&\Gamma+\frac{2\Gamma\|y_{k-1}^*\|\|d_{k-1}\|}{\bar{\gamma}^2}\leqslant\\&\Gamma+\frac{2\Gamma(L+t\Gamma^r)\alpha_{k-1}\|d_{k-1}\|}{\bar{\gamma}^2}\|d_{k-1}\|\end{aligned} \tag{19}$$

由式(6)和假设(A)知

$$\alpha_{k-1}d_{k-1}\to 0(k\to\infty) \tag{20}$$

故存在一个常数 $\varepsilon\in(0,1)$ 和一个整数 k_0,对所有的 $k\geqslant k_0$ 有

$$\frac{2\Gamma(L+t\Gamma^r)}{\bar{\gamma}^2}\alpha_{k-1}\|d_{k-1}\|\leqslant\varepsilon \tag{21}$$

因此,对所有的 $k\geqslant k_0$,由式(19)和(21)有

$$\|d_k\|\leqslant\Gamma+\varepsilon\|d_{k-1}\|\leqslant\Gamma(1+\varepsilon+\varepsilon^2+\ldots+\varepsilon^{k-k_0-1})+\varepsilon^{k-k_0}\|d_{k_0}\|\leqslant\frac{\Gamma}{1-\varepsilon}+\|d_{k_0}\|$$

令 $M=\max\{\|d_1\|,\|d_2\|,\cdots,\|d_{k_0}\|,\frac{\Gamma}{1-\varepsilon}+\|d_{k_0}\|\}$,故 $\|d_k\|\leqslant M$.

定理 1　假设(A)和假设(B)成立,d_k 由式(11)确定,α_k 由 Armijo-type 线搜索获得,则 $\lim\limits_{k\to\infty}\inf\|g_k\|=0$.

证明　如果定理 1 不成立,则有式(16)成立.如果 $\lim\limits_{k\to\infty}\inf\alpha_k>0$,则由式(15)和(20)式有 $\lim\limits_{k\to\infty}\inf\|g_k\|=0$,这与假设矛盾.如果 $\lim\limits_{k\to\infty}\inf\alpha_k=0$,则存在无限指标集 K,使得

$$\lim_{k\in K,k\to\infty}\alpha_k = 0 \tag{22}$$

由算法步骤 2 可知,当 $k\in K$ 充分大时,$\rho^{-1}\alpha_k$ 不满足式(6),即

$$f(x_k+\rho^{-1}\alpha_k d_k)-f(x_k)>-\delta\rho^{-2}\alpha_k^2\|d_k\|^2 \tag{23}$$

由中值定理,引理(2),式(13)和(15),存在 $h_k\in(0,1)$,使得

$$\begin{aligned} f(x_k+\rho^{-1}\alpha_k d_k)-f(x_k)=\rho^{-1}\alpha_k g(x_k+h_k\rho^{-1}\alpha_k d_k)^{\mathrm{T}}d_k= \\ \rho^{-1}\alpha_k g_k^{\mathrm{T}}d_k+\rho^{-1}\alpha_k(g(x_k+h_k\rho^{-1}\alpha_k d_k)-g_k)^{\mathrm{T}}d_k\leqslant \\ \rho^{-1}\alpha_k g_k^{\mathrm{T}}d_k+L\rho^{-2}\alpha_k^2\|d_k\|^2 \end{aligned}$$

再由式(23),当 $k\in K$ 充分大时,有

$$\|g_k\|^2\leqslant\rho^{-1}(L+\delta)\alpha_k\|d_k\|^2$$

因为 $\{d_k\}$ 有界,$\lim_{k\in K,k\to\infty}\alpha_k=0$,则 $\lim_{k\to\infty}\inf\|g_k\|=0$.这与假设矛盾,故结论成立.

3 结 论

对于求解大规模无约束优化问题,给出了一种修正的三项 LS 共轭梯度法,该方法不依赖任何线搜索,具有充分下降性,并且在 Armijo-type 线搜索下证明了该方法的全局收敛性.

参考文献:

[1] ANDREI N. A Dai-Yuan Conjugate Gradient Algorithm with Sufficient Descent and Conjugacy Conditions for Unconstrained Optimization[J].Applied Mathematics Letters,2008,21(2):165-171

[2] LIU Y,STOREY C.Efficient Generalized Conjugate Gradient Algorithms,Part 1:Theory[J].Journal of Optimization Theory and Applications,1991,69(1):129-137

[3] GILBERT J C,NOCEDAL J.Global Convergence Properties of Conjugate Gradient Methods for Optimization[J].SIAM Journal on Optimization,1992,2(1):21-42

[4] YU G,GUAN L,LI G.Global Convergence of Modified Polak-Ribiere-Polyak Conjugate Gradient Methods with Sufficient Descent Property[J].Journal of Industrialand Management Optimization,2008(4):565-579

[5] ZHANG L,ZHOU W,LI D H.A Descent Modified Polak-Ribière-Polyak Conjugate Gradient Method and Its Global Convergence [J].IMA Journal of Numerical Analysis,2006,26(4):629-640

[6] LI D H, FUKUSHIMA M. A Modified BFGS Method and Its Global Convergence in Nonconvex Minimization[J]. Journal of Computational and Applied Mathematics,2001(1):15-35

[7] NARUSHINMA Y, YABE H, FIORD J A. A Three-term Conjugate Gradient Method with Sufficient Descent Property for Unconstrained Optimization[J].SIAM Journal on Optimization,2011,21(1):212-230

A Modified Three-term LS Conjugate Gradient mMethod

JIAO Jia-jia

(College of Mathematics Science, Chongqing Normal University, Chongqing 401331, China)

Abstract: Based on the existent ideas of the conjugate gradient method, this paper proposes a three-term LS conjugate gradient method. The presented method can make the search direction with sufficient descent property without any line search. Under certain mild conditions, global convergence is obtained.

Key words: Three-term Conjugate Gradient Method;sufficient descent property;global convergence