doi:10.16055/j.issn.1672-058X.2015.0001.001

一种充分下降的 CD 共轭梯度法及其收敛性*

马　烁, 王安平**
(荆州理工职业学院,湖北 荆州 434000;长江大学 工程技术学院,湖北 荆州 434020)

摘　要:基于已有的 CD 方法,提出了一种改进的 CD 共轭梯度法(MCD 算法).该算法产生的搜索方向为充分下降方向,且这一性质与所采用的线搜索方法无关;并在一定的条件下证明了该算法基于 Wolfe 线搜索求解非凸优化问题的全局收敛性.

关键词:无约束最优化;共轭梯度法;Wolfe 线搜索;全局收敛性

中图分类号:O29　　**文献标识码**:A　　**文章编号**:1672-058X(2015)01-0001-04

0　引　言

考虑无约束优化问题

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

其中 $f:\mathbf{R}^n \to \mathbf{R}$ 连续可微.共轭梯度法是求解该问题的一类有效算法.

一般的共轭梯度法的迭代公式为

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

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

其中,x_1 为初始点,d_k 为搜索方向,α_k 由某种线性搜索或由特定公式计算出的步长因子,β_k 为标量,$g_k = \nabla f(x_k)$.

共轭梯度法的关键是选取 α_k 和 β_k,不同的 α_k 和 β_k 决定了不同的共轭梯度算法.常用选取 α_k 的线搜索是标准 Wolfe 线搜索,即选取 $\alpha_k>0$ 满足

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

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

其中 δ 和 σ 是满足 $0<\delta<\sigma<1$ 的常数;而 β_k 的选取公式常用的有

$$\beta_k^{\mathrm{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \beta_k^{\mathrm{PRP}} = \frac{g_k^{\mathrm{T}}(g_k - g_{k-1})}{\|g_{k-1}\|^2}, \beta_k^{\mathrm{HS}} = \frac{g_k^{\mathrm{T}}(g_k - g_{k-1})}{(g_k - g_{k-1})^{\mathrm{T}} d_{k-1}}$$

$$\beta_k^{\mathrm{CD}} = -\frac{\|g_k\|^2}{g_{k-1}^{\mathrm{T}} d_{k-1}}, \beta^{\mathrm{L}}\mathrm{S}_k = \frac{g_k^{\mathrm{T}}(g_k - g_{k-1})}{g_{k-1}^{\mathrm{T}} d_{k-1}}, \beta_k^{\mathrm{DY}} = \frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}} y_{k-1}}$$

其中$\|\cdot\|$为欧式范数,它们对应的共轭梯度法称为 FR[1]方法、PRP[2]方法、HS[3]方法、CD[4]方法、LS[5]方法和 DY[6]方法.

收稿日期:2014-05-29;**修回日期**:2014-06-10.

* **基金项目**:湖北省教育科学“十二五”规划课题(2012B310);长江大学工程技术学院基金(13J0802).

作者简介:马烁(1981-),女,陕西蓝田人,讲师,硕士,从事最优化理论与算法研究.

** **通讯作者**:王安平(1980-),男,陕西旬阳人,讲师,硕士,从事最优化理论与算法研究.E-mail:wanganping2010@126.com.

在众多共轭梯度法中,为了保证是下降方向,许多学者都做了深入的研究.文献[7]提出一种新的谱 DY 共轭梯度法(简记为 MDY),其迭代格式为式(2)和

$$d_k = \begin{cases} -g_k, & \text{if } k = 1 \\ -\theta_k g_k + \beta_k d_{k-1}, & \text{if } k \geqslant 1 \end{cases}$$

其中 $\theta_k = 1 + \frac{g_k^{\mathrm{T}} d_{k-1}}{d_{k-1}^{\mathrm{T}} y_{k-1}}$,$\beta_k = \beta_k^{\mathrm{DY}}$.MDY 方法可以产生不依赖于线搜索的下降方向,且在 Armijo 型线搜索规则下具有全局收敛性.

在文献[8]中,Hager 和 Zhang 提出了一种新的共轭梯度法(简称 HZ 方法).HZ 方法的一个重要特性是它可以产生不依赖于线搜索方法的下降方向,且满足充分下降条件 $d_k^{\mathrm{T}} g_k \leqslant -\frac{7}{8}\|g_k\|$.Hager 和 Zhang 证明了采用 Wolfe 搜索时,HZ 算法的全局收敛性.

受到文献[8]的启发,结合 CD 方法,此处提出了一种修正的 CD 算法,该算法具有不依赖于所采用的线搜索方法的充分下降性;证明了该算法在 Wolfe 线搜索下的全局收敛性;通过数值试验验证了算法的有效性.

1 修正的 CD 算法

提出的修正 CD 共扼梯度法与 HZ 方法有相似的结构,称之为 MCD 方法.参数 β_k^{MCD} 的计算公式如下:

$$\beta_k^{\mathrm{MCD}} = \beta_k^{\mathrm{CD}} \left(1 - \frac{\mu g_k^{\mathrm{T}} d_{k-1}}{d_{k-1}^{\mathrm{T}} g_{k-1}}\right) \tag{6}$$

其中 $\mu > \frac{1}{4}$ 为一常数,由 β_k^{MCD} 的计算式(6),给出如下充分下降性定理.

定理 1 设迭代方向由式(7)产生

$$d_k = -g_k + \beta_k^{\mathrm{MCD}} d_{k-1}, d_0 = -g_0 \tag{7}$$

若 $d_{k-1}^{\mathrm{T}} y_{k-1} \neq 0$,则有

$$d_k^{\mathrm{T}} g_k \leqslant \left(\frac{1}{4\mu} - 1\right) \|g_k\|^2 \tag{8}$$

证明 当 $k=0$ 时,$d_0^{\mathrm{T}} g_0 = -\|g_0\|^2$,式(8)成立.

当 $k \geqslant 1$ 时,$d_k = -g_k + \beta_k^{\mathrm{MCD}} d_{k-1}$,两边与 g_k 做内积,得

$$g_k^{\mathrm{T}} d_k = -\|g_k\|^2 + \beta_k^{\mathrm{MCD}} g_k^{\mathrm{T}} d_{k-1} =$$

$$-\|g_k\|^2 + \frac{g_k^{\mathrm{T}} d_{k-1} \|g_k\|^2}{(-d_{k-1}^{\mathrm{T}} g_{k-1})}\left(1 - \frac{\mu g_k^{\mathrm{T}} d_{k-1}}{(-d_{k-1}^{\mathrm{T}} g_{k-1})}\right) =$$

$$\frac{\|g_k\|^2 (g_k^{\mathrm{T}} d_{k-1})(d_{k-1}^{\mathrm{T}} g_{k-1}) - \|g_k\|^2 (-d_{k-1}^{\mathrm{T}} g_{k-1})^2 - \mu \|g_k\|^2 (g_k^{\mathrm{T}} d_{k-1})^2}{(-d_{k-1}^{\mathrm{T}} g_{k-1})^2} \leqslant$$

$$\frac{\frac{1}{2}\left[2\mu \|g_k\|^2 (g_k^{\mathrm{T}} d_{k-1})^2 + \frac{1}{2\mu}\|g_k\|^2 (d_{k-1}^{\mathrm{T}} g_{k-1})^2\right] - \|g_k\|^2 (-d_{k-1}^{\mathrm{T}} g_{k-1})^2 - \mu \|g_k\|^2 (g_k^{T} d_{k-1})^2}{(-d_{k-1}^{\mathrm{T}} g_{k-1})^2} =$$

$$\left(\frac{1}{4\mu} - 1\right)\|g_k\|^2$$

故定理 1 得证.

由定理 1 知,由式(7)产生的迭代方向均为下降方向,且该下降性的产生不依赖于任何线性搜索.

算法 1 MCD 算法

步 1 给定初始点 $x_1 \in \mathbf{R}^n$,$\varepsilon > 0$,$d_1 = -g_1$,令 $k := 1$;

步 2 若 $\|g_k\| \leqslant \varepsilon$,则停止迭代,否则转入步 3;

步 3 由式(4)和式(5)求得 α_k;

步 4　计算 $x_{k+1}=x_k+\alpha_k d_k$,若$\|g_{k+1}\|\leqslant\varepsilon$,则算法停止,否则转步 5;

步 5　利用式(6)计算 β_{k+1},计算 $d_{k+1}=-g_{k+1}+\beta_{k+1}d_k$,置 $k:=k+1$,转步 2.

2　算法的全局收敛性

下面,将在一定的假设条件下证明 MCD 算法的全局收敛性,如无特殊说明,下面的假设 A 总成立.

假设 A

1) 水平集 $L_1=\{x\in\mathbf{R}^n \mid f(x)\leqslant f(x_1)\}$有界,其中 x_1 为初始点;

2) 在水平集 L_1 的一个邻域 U 内,$f(x)$是连续可微的,其梯度 $g(x)$是 lipschitz 连续的,即存在常数 $L>0$,使

$$\|g(x)-g(y)\|\leqslant L\|x-y\|,\ \forall x,y\in U \tag{9}$$

引理 1　设目标函数 $f(x)$满足假设 A,序列$\{x_k\}$由式(2)和式(3)产生,其中 β_k 由式(6)计算,α_k 满足式(4)(5),则$\sum\limits_{k=1}^{\infty}\dfrac{\|g_k\|^4}{\|d_k\|^2}<+\infty$,此关系式称为 Zoutendijk 条件.

证明　由定理 1 及式(8)(9),可得

$$(\sigma-1)g_k^{\mathrm{T}}d_k\leqslant(g_{k+1}-g_k)^{\mathrm{T}}d_k\leqslant L\alpha_k\|d_k\|^2$$

因此,

$$\alpha_k\geqslant\frac{(\sigma-1)}{L}\frac{g_k^{\mathrm{T}}d_k}{\|d_k\|^2}=\frac{(1-\sigma)}{L}\frac{|g_k^{\mathrm{T}}d_k|}{\|d_k\|^2} \tag{10}$$

式(10)结合式(8)可得

$$\alpha_k\geqslant c\frac{\|g_k\|^2}{\|d_k\|^2},c=\frac{(1-\sigma)(4\mu-1)}{4L\mu},\mu>\frac{1}{4} \tag{11}$$

由式(4)和假设 A,有

$$-\sum_{k=1}^{\infty}\alpha_k g_k^{\mathrm{T}}d_k<+\infty \tag{12}$$

结合式(8)(12)和(13)可得条件 Zoutendijk 成立,定理 1 得证.

定理 2　设目标函数 $f(x)$满足假设条件 A,序列$\{x_k\}$由式(2)与式(3)产生,其中 β_k 由式(6)计算,α_k 由式(4)(5)确定,则 $g_k=0$ 对于某些指标 k 成立,或 $\liminf\limits_{k\to\infty}\|g_k\|=0$.

证明　类似于文献[9],此处不再赘述.

3　数值试验

比较 MCD 与标准 CD 方法的性能,两种方法均采用 Wolfe 线搜索,试验中的函数全部来源于文献[9].测试环境为 MATLAB 7.9.0,运行环境为内存 4.00 GB,CPU 为 2.60 GHz 的个人计算机.参数设置如下:$\delta=0.35$,$\sigma=0.75$,$0.25<\mu<0.5$;终止条件为$\|g_k\|\leqslant 10^{-6}$或 It-limit>9 999.计算得到 2 种方法的数值结果如表 1 所示.

表 1　方法 CD 与 MCD 数值测试结果

函数名称	维数	MCD	CD
		NI/NF/NG/T	NI/NF/NG/T
Rosenbrock	2	253/380/311/8.403 0	944/1 245/2 069/19.3
Freudenstein and roth	2	34/67/45/1.968 7	F
	8	128/191/151/31.713 8	210/287/244/33.170 4
Beale	2	45/87/60/3.162 9	F
	8	143/208/174/41.367 9	F
Bard	3	4/10/7/0.397 9	4/10/7/1.679 3
Gaussian	3	4/9/5/0.456 1	4/9/5/1.513 5
Box three-dimensional	3	23/51/33/5.380 5	118/165/134/6.522 6

续表1

函数名称	维数	MCD	CD
		NI/NF/NG/T	NI/NF/NG/T
Extendend Powell singular	4	489/605/534/12.216 6	546/677/585/22.679 4
	20	587/722/621/232.768 6	729/874/767/503.466 5
Brown and Dennis	4	198/260/213/393.829 7	F
Osborne 1	5	10/23/14/4.793 3	9/34/13/14.312 7
Extended miele and casntrell	4	361/504/426/12.612 1	556/758/652/83.803 5
Raydan 1	10	55/90/61/4.695 6	29/58/32/10.057 3
Raydan 2	10	5/11/8/3.993 0	6/13/9/1.818 2
	100	6/14/9/20.186 9	6/14/9/23.383 8
Generalized iridiagonal 1	10	134/173/143/20.466 9	166/220/184/32.191 0

注:“NI/NF/NG/Time”依次表示迭代的次数、函数值计算的次数、梯度值计算的次数和 CPU 运行总时间,“F”表示该方法失败.

通过以上简单数值试验,说明了新方法的可行性,但为了能得到更好的数值效果,还需要进一步研究方法中参数的选取和搜索条件的选择.

参考文献:

[1] FLETCHER R,REEVES C.Function Minimization by Conjugate Gradients[J].Computer Journal,1964(7):149-154
[2] POLAK E,RIBIERE G.Note Sur La Convergence De Dirctions Conjugees[J].Rev Fran-caise Informat Recherche Opertionelle,3e Annee,1969(16):35-43
[3] HESTEMES M R,SRIEFEL E L.Methods of Conjugate Gradient for Solving Linear Systems[J].Journal of Research of the National Bureau of Standards,1952,6(49):40-43
[4] FLETCHER R.Practical Methods Optimization[C]//Unconstrained optimization.New York:John Wiley&sons,1987
[5] LIU Y,STOREY C.Effcient Generalized Conjugate Gradient Algorithms[J].Journal of Optimiztion Theory and Applicatons,1991(69):129-137
[6] DAI Y H,YUAN Y.A Nonlinear Conjugate Gradient with a Strong Global Conver-gence Property[J].SIAM Journal on Optimizton,2000(10):177-182
[7] 敖卫斌.一种修正的 DY 共轭梯度法的全局收敛性[J].重庆工商大学学报:自然科学版,2013,30(10):17-20
[8] HAGER WILLIAM W,ZHANG HCH.A New Conjugate Gradient Method with Guaranteed Descent and an Efficient Line Search[J].SIAM Journal on Optimization ,2005,16(1):170- 92
[9] MORE J J,GARBOW B S,HILLSTROME K E.Testing Unconstrained Optimization Software[J].ACM Trans Math sofetware,1981(7):17-41

A Sufficiently Descent CD Conjugate Gradient Method and Its Convergence

MA Shuo[1] ,WANG An-ping[2]

(1. Jinzhou Vocational College of Technology,Hubei Jinzhou 434000,China;
2. School of Engineering and Technology,Yangtze River University,Hubei Jinzhou 434020,China)

Abstract: Based on CD method, this paper proposes a modified CD conjugate gradient method (MCD method), the search direction generated by this algorithm is sufficiently descent direction and this property is nothing to do with the used line search method,and under certain condition,the solution to the global convergence of non-convex-optimization problem based on Wolfe line searching is proved.

Key words: unconstrained optimality;conjugate gradient method;Wolfe line searching;global convergence

责任编辑:李翠薇