DOI:10.13656/j.cnki.gxkx.2013.04.020

广西科学 Guangxi Sciences 2013,20(4):348～351

一个三项 LS 共轭梯度方法*

A Three-Term LS Conjugate Gradient Method

李向荣

LI Xiang-rong

(**广西大学数学与信息科学学院,广西南宁** 530004)

(College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, 530004, China)

摘要:给出一个三项 LS 共轭梯度方法,并在适当条件下获得此方法对一般函数的全局收敛性.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,而且比通常的 LS 方法更具竞争性.

关键词:三项共轭梯度　充分下降　全局收敛性

中图法分类号:O110.74　　**文献标识码**:A　　**文章编号**:1005-9164(2013)04-0348-04

Abstract: A three-term LS conjugate gradient method is proposed, which can ensure that the search direction possesses the sufficient descent property without any line search. Under suitable conditions, the global convergence of the general function will be established. Moreover, numerical results show that the new method is more competitive to the usual method.

Key words: three-term conjugate gradient, sufficient descent, global convergence

对于问题

$$\min f(x)$$

$$x \in R^n, \tag{0.1}$$

其中 $f(x)$ 是 $R^n \to R$ 的连续可微函数,有很多求解方法.其中共轭梯度法结构简单且效果明显,特别是针对大规模优化问题.该方法已是优化界的研究热点,其一般迭代公式为

$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1,2,\cdots$$

x_k 是第 k 次迭代点,$\alpha_k > 0$ 是步长,d_k 是搜索方向,其定义形式为

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

其中 $\beta_k \in Re$ 是一个参数.人们根据 β_k 的选取不同来命名不同的共轭梯度法.著名的共轭梯度法(PRP)中 β_k 的定义式[1,2]为

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

其中 $g_k = \nabla f(x_k)$,$g_{k+1} = \nabla f(x_{k+1})$ 分别表示函数 $f(x)$ 在 x_k 和 x_{k+1} 的梯度值,$\| \cdot \|$ 是欧氏向量范数.PRP 方法具有自动再开始的性质,数值表现优越.但是 PRP 方法的收敛性不理想,Powell[13] 曾列举过一个反例说明 PRP 方法对某些非凸函数不具有全局收敛性,原因是参数 β_k^{PRP} 可能小于零.鉴于此,Gilbert 等[4] 证明 $\beta_k^{+} = \max\{0, \beta_k^{\mathrm{PRP}}\}$ 方法在弱 Wolfe-Powell 线搜索下的全局收敛性.还有研究已发现,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 k,下式成立

$$d_k^T g_k \leqslant -c \| g_k \|^2. \tag{0.2}$$

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[5~11].如 Yuan[7] 给出一个修改的 PRP 公式:

$$\beta_k^{\mathrm{MPRP}} = \beta_k^{\mathrm{PRP}} - \min\{\beta_k^{\mathrm{PRP}}, \frac{\mu \| y_k \|^2}{\| g_k \|^4} g_{k+1}^T d_k\},$$

其中 $\mu > \frac{1}{4}$ 是常数,$y_k = g_{k+1} - g_k$.该方法能保证 $\beta_k^{\mathrm{MPRP}} \geqslant 0$,且自动拥有充分下降性,所以其数值表现优于 PRP 方法.Zhang 等[12] 给出一个 PRP 三项共轭梯度方法,其方向定义为

收稿日期:2013-02-26

修回日期:2013-05-10

作者简介:李向荣(1979-),女,讲师,主要从事优化方法及其在金融模型的应用研究。

*国家自然科学基金项目(编号:11261006),广西大学科研基金项目(编号:XJZ110632),广西高等学校特色专业及课程一体化建设项目(管理科学专业)和广西自然科学基金项目(编号:2012GXNSFAA053002)资助。

$$d_{k+1}=\begin{cases}-g_{k+1}+\beta_k^{PRP}d_k-\vartheta_k y_k, \text{if}, k\geqslant 1,\\ -g_{k+1}, \text{if}, k=0,\end{cases}$$

其中 $\vartheta_k=\frac{g_{k+1}^T d_k}{\|g_k\|^2}$，$\beta_k^{PRP}=\frac{g_{k+1}^T y_k}{\|g_k\|^2}$，容易得到 $d_k^T g_k=-\|g_k\|^2$. 而通常的 LS 公式的 β_k 定义[13] 如下

$$\beta_k^{LS}=\frac{g_{k+1}^T y_k}{-d_k^T g_k}.$$

与三项 PRP 方法类似，具有充分下降性的三项 LS 公式定义为

$$d_{k+1}=\begin{cases}-g_{k+1}+\beta_k^{LS}d_k+\frac{d_k^T g_{k+1}}{d_k^T g_k}y_k, \text{if}, k\geqslant 1,\\ -g_{k+1}, \text{if}, k=0.\end{cases} \quad (0.3)$$

此方法仅仅利用梯度值信息，忽略了函数值信息. 最近文献[14] 给出了新的拟牛顿方程

$$B_{k+1}s_k=y_k^*,$$

其中 $y_k^*=y_k+\gamma_k s_k$，$\gamma_k=\max\{0, \frac{(g_{k+1}+g_k)^T s_k+3(f(x_k)-f(x_{k+1}))}{\|s_k\|^2}\}$，$B_{k+1}$ 是拟牛顿矩阵. 此拟牛顿方程中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，本文给出一个三项 LS 共轭梯度方法

$$d_{k+1}=\begin{cases}-g_{k+1}+\beta_k^{MLS}d_k+\frac{d_k^T g_{k+1}}{d_k^T g_k}y_k^*, \text{if}, k\geqslant 1,\\ -g_{k+1}, \text{if}, k=0,\end{cases} \quad (0.4)$$

其中 $\beta_k^{MLS}=\frac{g_{k+1}^T y_k^*}{-d_k^T g_k}$. 可以看出新公式不但拥有梯度值还拥有函数值信息.

1 修改的三项 LS 共轭梯度算法

为表述需要，称修改的三项 LS 共轭梯度算法为算法 1，其步骤如下：

步骤 0　给定 $x_0\in Re^n$，$\delta\in(0,1/2)$，$\sigma\in(\delta,1)$ 和终止参数 $\varepsilon>0$. 令 $d_0=-g_0=-\nabla f(x_0)$，置 $k:=0$.

步骤 1　若 $\|g_k\|\leqslant\varepsilon$，停止.

步骤 2　寻找满足下面 WWP 线搜索的步长 α_k，

$$f(x_k+\alpha_k d_k)\leqslant f_k+\delta\alpha_k g_k^T d_k, \quad (1.1)$$

$$g(x_k+\alpha_k d_k)^T d_k\geqslant\sigma g_k^T d_k. \quad (1.2)$$

步骤 3　令 $x_{k+1}=x_k+\alpha_k d_k$. 如果 $\|g_{k+1}\|\leqslant\varepsilon$，停止.

步骤 4　利用公式(0.4) 计算搜索方向.

步骤 5　置 $k:=k+1$，转步骤 2.

2 算法 1 的充分下降性和全局收敛性

引理 2.1　对 $k\geqslant 0$，三项 LS 公式的搜索方向满足

$$d_k^T g_k=-\|g_k\|^2. \quad (2.1)$$

证明　如果 $k=0$，则 $g_0^T d_0=-\|g_0\|^2$，那么 (2.1) 式成立. 当 $k\geqslant 1$ 时，利用新方向的定义可得

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

证毕.

假设有如下条件(A)：

(i) 水平集 $\Omega=\{x\in R^n: f(x)\leqslant f(x_0)\}$ 有界.

(ii) f 在 Ω 上有下界且连续可微，它的梯度 g 满足 Lipschitz 条件，即存在常数 $L>0$ 满足

$$\|g(x)-g(y)\|\leqslant L\|x-y\|, \forall, x, y\in\Omega. \quad (2.2)$$

定理 2.1　设序列$\{g_k\}$和$\{d_k\}$由算法1产生，若假设条件(A) 满足，且存在常数 $\rho>0$ 使得 $\|d_k\|\leqslant\rho\|g_k\|$ 成立，则 $\lim\limits_{k\to\infty}\|g_k\|=0$ 成立.

证明　由(1.2) 式和 Lipschitz 条件(2.2)，得

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

又由(2.1) 式，得

$$\alpha_k\geqslant\frac{(1-\sigma)}{L}\frac{|g_k^T d_k|}{\|d_k\|^2},$$

把上式代入(1.1) 式并移向得

$$\frac{\delta(1-\sigma)}{L}\frac{(g_k^T d_k)^2}{\|d_k\|^2}\leqslant f_k-f_{k+1}.$$

取 $k=0$ 到 ∞ 将上式相加，得到

$$\sum_{k=0}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2}\leqslant\frac{L}{\delta(1-\sigma)}(f_0-f_\infty).$$

由假设条件(A) 中的函数 f 有下界，有

$$\sum_{k=0}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2}<+\infty. \quad (2.3)$$

又由充分下降性(2.1) 式和 $\|d_k\|\leqslant\rho\|g_k\|$，可得 $\sum\limits_{k=0}^{\infty}\|g_k\|^2<+\infty$，所以 $\|g_k\|\to 0$ 成立，证毕.

3 数值检验

选择工程中常用的 Benchmark 问题(http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume24/ortizboyer05a-html/node6.html). 列举如下：

(1) $f_{Sph}(x)=\sum\limits_{i=1}^{n}x_i^2$，$x_i\in[-5.12, 5.12]$，$x^*=(0,0,\cdots,0)$，$f_{Sph}(x^*)=0$.

(2) $f_{SchDS}(x)=\sum\limits_{i=1}^{n}(\sum\limits_{j=1}^{i}x_j)^2$，$x_i\in[-65.536,$

65.536]$,x^{*}=(0,0,\cdots,0),f_{SchDS}(x^{*})=0$.

(3)$f_{\mathrm{Ras}}(x)=10n+\sum_{i=1}^{n}(x_i^{2}-10\cos(2\pi x_i)),x_i\in[-5.12,5.12],x^{*}=(0,0,\cdots,0),f_{Ras}(x^{*})=0$.

(4)$f_{\mathrm{Gri}}(x)=1+\sum_{i=1}^{n}\frac{x_i^2}{4000}-\prod_{i=1}^{n}\cos\frac{x_i}{i},x_i\in[-600,600],x^{*}=(0,0,\cdots,0),f_{Gri}(x^{*})=0$.

终止条件：当 $|f(x_k)|>e_2$，令 stop $=\frac{|f(x_k)-f(x_{k+1})|}{f(x_k)}$，否则令 stop $=|f(x_k)-f(x_{k+1})|$. 若 $\|g(x_k)\|\leqslant\varepsilon$ 或 stop $\leqslant\varepsilon$，程序结束，$\varepsilon=10^{-5}$. 参数取值：$\delta=0.01,\sigma=0.9$. 表 1～2 中的各参数含义：P 是问题名；Dim 是问题维数；NI 是迭代次数；NFG 是函数和梯度次数和；$f(\bar{x})$ 是终止时函数值.

表 2 中通常的 LS 方法的搜索方向由下式定义

$$d_k=\begin{cases}-g_k,\text{if},k=0,\\-g_k+\beta_k^{\mathrm{LS}}d_{k-1},\text{if},k\geqslant 1.\end{cases}$$

从表 1 和表 2 的结果可以看出，两种算法的表现都比较理想，维数和初始点对 NI 和 NFG 的影响不算太大. 另外最终函数值非常接近最优值，有些已完全达到最优函数值. 两种算法的 NI 和最终函数值几乎一样，而算法 1 需要的 NFG 更少一些. 所以算法 1 对求解上述问题效果更具竞争性.

表 1 算法 1 的数值结果

Table 1 Numerical results of algorithm 1

P		NI/NFG/f $(\bar{x})$	NI/NFG/ $f(\bar{x})$	NI/NFG / $f(\bar{x})$	NI/NFG/ $f(\bar{x})$
Sphere	x_0	$(-4,\cdots,-4)$	$(4,\cdots,4)$	$(-4,0,-4,0,\cdots)$	$(4,0,4,0,\cdots)$
Dim	30	2/6/8.519698e−028	2/6/8.519698e−028	2/6/1.064962e−028	2/6/1.064962e−028
	100	2/6/1.262177e−027	2/6/1.262177e−027	2/6/4.831773e−028	2/6/4.831773e−028
	300	2/6/2.366583e−026	2/6/2.366583e−026	2/6/1.848893e−026	2/6/1.848893e−026
Schwefel's	x_0	$(-0.001,\cdots,-0.001)$	$(0.0001,\cdots,0.0001)$	$(-0.001,0,-0.001,0,\cdots)$	$(0.0001,0,0.0001,0,\cdots)$
Dim	30	5/14/4.584481e−007	3/8/1.017550e−007	4/11/2.183462e−006	3/8/4.480619e−008
	100	7/20/2.554559e−006	4/11/6.963498e−007	7/20/6.523845e−006	3/8/1.007628e−006
	300	12/35/4.895339e−006	6/17/1.766000e−006	24/71/8.607008e−006	5/14/1.454140e−006
Rastrigin	x_0	$(0.01,\cdots,0.01)$	$(0.001,\cdots,0.001)$	$(0.01,0,0.01,0,\cdots)$	$(0.001,0,0.001,0,\cdots)$
Dim	30	4/12/1.759304e−010	3/9/0.000000e+000	5/16/3.593324e−007	4/12/0.000000e+000
	100	4/12/0.000000e+000	4/11/0.000000e+000	4/12/6.116352e−011	4/11/6.693881e−010
	300	4/11/9.094947e−013	3/8/1.737135e−010	3/9/0.000000e+000	4/11/0.000000e+000
Griewank	x_0	$(-30,\cdots,-30)$	$(10,\cdots,10)$	$(-30,0,-30,0,\cdots)$	$(10,0,10,0,\cdots)$
Dim	30	2/6/0.000000e+000	24/89/9.864874e−003	2/6/0.000000e+000	6/33/0.000000e+000
	100	2/6/0.000000e+000	11/32/3.901090e−006	2/6/0.000000e+000	21/59/4.782356e−006
	300	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000	2/6/0.000000e+000

表 2 通常的 LS 算法的数值结果

Table 2 Numerical results of usual LS algorithm

P		NI/NFG/ $f(\bar{x})$	NI/NFG/ $f(\bar{x})$	NI/NFG/ $f(\bar{x})$	NI/NFG/ $f(\bar{x})$
Sphere	x_0	$(-4,\cdots,-4)$	$(4,\cdots,4)$	$(-4,0,-4,0,\cdots)$	$(4,0,4,0,\cdots)$
Dim	30	2/8/8.519698e−028	2/8/8.519698e−028	2/8/1.064962e−028	2/8/1.064962e−028
	100	2/8/1.262177e−027	2/8/1.262177e−027	2/8/4.831773e−028	2/8/4.831773e−028
	300	2/8/2.366583e−026	2/8/2.366583e−026	2/8/1.848893e−026	2/8/1.848893e−026
Schwefel's	x_0	$(-0.001,\cdots,-0.001)$	$(0.0001,\cdots,0.0001)$	$(-0.001,0,-0.001,0,\cdots)$	$(0.0001,0,0.0001,0,\cdots)$
Dim	30	5/22/4.584481e−007	3/12/1.017550e−007	4/17/2.183462e−006	3/12/4.480619e−008
	100	7/32/2.554559e−006	4/17/6.963498e−007	7/32/6.523845e−006	3/12/1.007628e−006
	300	12/57/4.895318e−006	6/27/1.766000e−006	23/112/8.615258e−006	5/22/1.454140e−006
Rastrigin	x_0	$(0.01,\cdots,0.01)$	$(0.001,\cdots,0.001)$	$(0.01,0,0.01,0,\cdots)$	$(0.001,0,0.001,0,\cdots)$
Dim	30	3/14/0.000000e+000	3/14/0.000000e+000	3/15/0.000000e+000	3/14/0.000000e+000
	100	3/14/0.000000e+000	3/14/0.000000e+000	3/14/0.000000e+000	3/14/0.000000e+000
	300	3/14/0.000000e+000	3/14/0.000000e+000	3/14/0.000000e+000	3/14/0.000000e+000
Griewank	x_0	$(-30,\cdots,-30)$	$(10,\cdots,10)$	$(-30,0,-30,0,\cdots)$	$(10,0,10,0,\cdots)$
Dim	30	2/8/0.000000e+000	12/70/2.836592e−006	2/8/0.000000e+000	6/43/0.000000e+000
	100	2/8/0.000000e+000	8/50/3.420712e−005	2/8/0.000000e+000	25/119/4.291810e−006
	300	2/8/0.000000e+000	2/8/0.000000e+000	2/8/0.000000e+000	2/8/0.000000e+000

致谢：

感谢广西大学数学与信息科学学院的韦增欣教授和袁功林教授给予的指导和帮助.

参考文献：

[1] Polak E,Ribiere G. Note sur la xonvergence dedirections conjugees[J]. Rev Francaise Informat Recherche Operatinelle,3e Annee,1969(16):35-43.

[2] Polyak B T. The conjugate gradient method in extreme problems[J]. USSR Comp Math Math Phys,1969(9):94-112.

[3] Powell M J D. Nonconvex minimization calculations and the conjugate gradient method,lecture notes in mathematics[J]. Spinger-Verlag,Berlin,1984,1066:122-141.

[4] Gilbert J C,Nocedal J. Global convergence properties of conjugate gradientmethods for optimization[J]. SIAM Journal on Optimization,1992,2:21-42.

[5] Hager W W,Zhang H. Algorithm 851:CG_DESCENT, A conjugate gradient method with guaranteed descent[J]. ACM Transactions on Mathematical Software,2006,32:113-137.

[6] Wei Z,Li G,Qi L. Global convergence of the PRP conjugate gradient methods with inexact line search for nonconvex unconstrained optimization problems[J]. Mathematics of Computation,2008,77:2173-2193.

[7] Yuan G L. Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems[J]. Optimization Letters,2009(3):11-21.

[8] Yuan G L,Lu X W. A modified PRP conjugate gradient method[J]. Annals of Operations Research,2009,166:73-90.

[9] Yuan G L,Lu X W,Wei Z X. A conjugate gradientmethod with descent direction for unconstrained optimization [J]. Journal of Computational and Applied Mathematics,2009,233:519-530.

[10] **赵许培,杨英芝,袁功林. 一个新的修正共轭梯度算法**[J]. **广西科学**,2012,19:104-107.

[11] Wei Z,Li G,Qi L. New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems[J]. Applied Mathematics and Computation,2006,179:407-430.

[12] Zhang L,Zhou W,Li D. A descent modified Polak-Ribiére-Polyak conjugate method and its global convergence[J]. IMA Journal on Numerical Analysis,2006,26:629-649.

[13] Liu Y,Storey C. Effcient generalized conjugate gradient algorithms,part 1: theory[J]. Journal of Optimization Theory and Application,1992,69:17-41.

[14] Yuan G,Wei Z. Convergence analysis of a modified BFGS method on convex minimizations[J]. Computational Optimization and Applications,2010,47:237-255.

（责任编辑:尹　闯）

广西修订科学技术奖励办法实施细则

新闻时间 2013-09-18

记者近日从广西科技厅获悉,为进一步规范管理区内科学技术奖励工作,保证自治区科学技术奖的质量,充分调动广大科技工作者的积极性和创造性,加速广西科学技术事业发展,科技厅对印发于 2010 年 10 月 29 日的《广西壮族自治区科学技术奖励办法实施细则(试行)》进行修订。此次修订内容包括进一步优化奖励结构、扩大项目公示内容,实行更严格的防止搭车报奖措施,实行每年聘请 1 次评审委员会的制度,落实更严格的回避措施,以此做好未来 3 年的科技奖励工作。据了解,参加此次修订的组成人员来自政府部门、科研院所、高等院校以及生产一线企业,有着较强的代表性、权威性和科学性。

据悉,《广西壮族自治区科学技术奖励办法实施细则(修订)》日前已印发至区内各市科技局,区直有关部门(单位)和中直驻桂有关单位。

摘自广西新闻网