采用广义Curry线搜索的重新开始共轭梯度算法及其全局收敛性*

焦宝聪　　陈兰平

(首都师范大学数学系，北京100037)

摘　要　本文对无约束最优化问题：$\min f(x), x\in R^n$，提出一种新的重新开始共轭梯度算法．该算法采用一类广义Curry线搜索原则，参数β_k可在一个有限闭区间内选择，且允许β_k取负值．在较弱的条件下证明了该算法的全局收敛性．

关键词　重新开始共轭梯度法，广义Curry线搜索，全局收敛性

分类号　AMS(1991) 90C30/CCL O221.2

1　引　言

考虑无约束最优化问题

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

其中函数$f: R^n\to R^1$连续可微．众所周知，由于共轭梯度算法不需存储迭代矩阵，并且又具有收敛速度较快和二次终止性的优点，因此，当决策变量个数n比较大时，是求解问题(1)的有效算法．这类算法的一般模式是：

Step 1．取初始点x_1，置$k:=1$．

Step 2．计算$g_k=\nabla f(x_k)$．

Step 3．若$g_k=0$，停止计算；否则置$P_k=-g_k+\beta_{k-1}P_{k-1}$，其中$\beta_{k-1}$按某种方式确定，$P_0=0$．

Step 4．一维线搜索，计算正步长α_k．

Step 5．令$x_{k+1}=x_k+\alpha_kP_k$，置$k:=k+1$，转Step 2．

最著名的共轭梯度法是Fletcher和Reeves(1964年)[1]提出的，通常称为FR共轭梯度法；由Polak，Ribiere[2]和Polyak(1969)提出的，通常称为PRP共轭梯度法．二者确定参数β_{k-1}的公式分别为

$$\beta_{k-1}^{FR}=\|g_k\|^2/\|g_{k-1}\|^2, \tag{2}$$

$$\beta_{k-1}^{PRP}=g_k^T(g_k-g_{k-1})/\|g_{k-1}\|^2. \tag{3}$$

关于FR共轭梯度法对一般非凸函数的全局收敛性，采用精确线搜索的情况由Powell(1983)[3]证明；采用非精确线搜索的情况由Al-Baali(1985年)[4]证明．在$f(x)$强凸假设下，

---

* 1996年7月17日收到

PRP 共轭梯度法在精确/非精确线搜索下的全局收敛性由袁亚湘(1993)[5]得到 尽管在有些情况下, PRP 算法可能比 FR 算法有效, 但 Powell(1984年)[6]给出的反例说明: 即使目标函数 $f(x)$ 二阶连续可微, 水平集 $\{x \in R^n, f(x) \le f(x_1)\}$ 有界, 并且采用精确线搜索, PRP 算法也可能不收敛 Powell 建议应限制 $\beta_k \ge 0$. 但 Grippo 和 Lucidi(1995年)[7]论证了选择 $\beta_k \ge 0$ 可能不是使 PRP 算法全局收敛的唯一途径 因此应当进一步研究共轭梯度算法中关于参数 β_k 的选择策略

本文给出了 β_k 可以取负值的条件, 以及 β_k 的取值范围; 并给出了采用一类广义 Curry 线搜索策略重新开始的共轭梯度法, 证明了算法的全局收敛性

2　定义与假设

假设(i) 水平集 $L_1 = \{x \mid x \in R^n, f(x) \le f(x_1)\}$ 有界;

(ii) 在 L_1 上, 目标函数 $f(x)$ 连续可微, 且其梯度 $g(x) = \nabla f(x)$ 有界, 即存在常数 $L > 0$, 使任意 $x \in L_1$, 有 $\|g(x)\| \le L$.

定义1　若纯量函数 $\delta: [0, +\infty) \to [0, +\infty)$ 满足: 对任意数列 $\{t_k\} \subset [0, +\infty)$, 若 $\lim_{k\to\infty} \delta(t_k) = 0$, 则有 $\lim_{k\to\infty} t_k = 0$, 则称 $\delta(\cdot)$ 为强迫函数

定义2　在假设(i), (ii) 下, 取

$$\alpha = \sup\{\|g(x) - g(y)\| \mid x, y \in L_1\} > 0.$$

令

$$S(t) = \begin{cases} \inf\{\|x - y\|, x, y \in L_1 \text{且} \|g(x) - g(y)\| \ge t\}, & t \in [0, \alpha) \text{时}, \\ \lim_{\tau\to 0^+} S(\tau), & \text{当 } t \in [\alpha, +\infty) \text{时}, \end{cases}$$

则称 $S(t)$ 为 $g(x)$ 在 L_1 上的连续性逆模

定义3　确定步长的广义 Curry 原则是指: 在 x_k 处沿搜索方向 P_k, 选取步长 α_k, 使

$$\alpha_k = \min\{\alpha \mid g(x_k + \alpha P_k)^T P_k = \mu g(x_k)^T P_k, \alpha > 0\},$$

且满足条件

$$\left| g(x_k + \alpha_k P_k)^T P_k \right| \le -\sigma g_k^T P_k.$$

其中 $\mu \in [0, \sigma)$, $\sigma \in (0, \frac{2}{5}]$. 这里 $g_k = g(x_k)$.

3　主要结果

现在给出重新开始的改进 PRP 算法如下:

算法A

Step 1. 任取初始点 $x_1 \in R^n$, 置 $k := 1$.

Step 2. 计算 $g_k = g(x_k)$.

Step 3. 若 $g_k = 0$, 停止计算; 否则, 置

$$P_k = -g_k + \beta_{k-1} P_{k-1}, \tag{4}$$

其中,

$$\beta_{k-1}=\begin{cases}0, \text{当 } k-1 \text{是 } n \text{ 的整数倍时};\\ \beta^{(k-1)}, \text{ 其它}\end{cases} \tag{5}$$

这里

$$-\frac{\bar{\sigma}}{\sigma}\beta_{k-1}^{PRP}\le\beta^{(k-1)}\le\frac{\bar{\sigma}}{\sigma}\beta_{k-1}^{PRP}, \bar{\sigma}\in(0, \frac{2}{5}), \sigma\in(0, \frac{2}{5}], \tag{6}$$

而

$$\beta_{k-1}^{PRP}=\frac{g_k^T(g_k-g_{k-1})}{\|g_{k-1}\|^2}. \tag{7}$$

Step 4. 按广义 Curry 原则, 计算步长 α_k.

Step 5. 置 $x_{k+1}=x_k+\alpha_k P_k$, 计算 $g_{k+1}=g(x_{k+1})$.

若 $|g_{k+1}^T g_k|<0.2\|g_{k+1}\|^2$, 令 $k:=k+1$, 转 Step 3; 否则, 令 $P_k=-g_k$, 转 Step 4.

引理1　对算法A, 不等式

$$-\sum_{j=0}^{k-1}\bar{\sigma}^j\le\frac{g_k^T P_k}{\|g_k\|^2}\le-2+\sum_{j=0}^{k-1}\bar{\sigma}^j, \tag{8}$$

$$g_k^T P_k<0, \tag{9}$$

对所有 k 成立, 只要 $g_k\ne0$. 其中 $\bar{\sigma}\in(0, \frac{2}{5})$.

证明　用归纳法, 对 $k=1$, 因 $P_1=-g_1$, $\bar{\sigma}^0=1$, 故不等式(8), (9) 显然成立.

现假设对任何 $k\ge1$ 有(8) 式成立, 则由于

$$\sum_{j=0}^{k-1}\bar{\sigma}^j<\sum_{j=0}^{\infty}\bar{\sigma}^j=\frac{1}{1-\bar{\sigma}}, \tag{10}$$

(8) 式右端为负, 故此时(9) 式也成立.

由于

$$\frac{g_{k+1}^T P_{k+1}}{\|g_{k+1}\|^2}=\frac{g_{k+1}^T(-g_{k+1}+\beta_k P_k)}{\|g_{k+1}\|^2}=-1+\beta_k\frac{g_{k+1}^T P_k}{\|g_{k+1}\|^2},$$

当 $\beta_k=0$ 或 $|g_{k+1}^T g_k|\ge0.2\|g_{k+1}\|^2$ 时, (8)、(9) 两个不等式显然成立. 考虑 $\beta_k=\beta^{(k)}$ 且 $|g_{k+1}^T g_k|<0.2\|g_{k+1}\|^2$. 此时, 由(6), (7) 式及定义3中 α_k 的取法, 有

$$\begin{aligned}\frac{g_{k+1}^T P_{k+1}}{\|g_{k+1}\|^2}&\le-1+\mu\frac{\bar{\sigma}}{\sigma}\frac{g_{k+1}^T(g_{k+1}-g_k)}{\|g_{k+1}\|^2}\cdot\frac{g_k^T P_k}{\|g_k\|^2}\le-1+1.2\bar{\sigma}(-2+\sum_{j=0}^{k-1}\bar{\sigma}^j)\\&\le-1+\bar{\sigma}(-2+\sum_{j=0}^{k-1}\bar{\sigma}^j)\le-1+\bar{\sigma}\sum_{j=0}^{k-1}\bar{\sigma}^j=-2+\sum_{j=0}^{k}\bar{\sigma}^j.\end{aligned}$$

同理, 有

$$\begin{aligned}\frac{g_{k+1}^T P_{k+1}}{\|g_{k+1}\|^2}&\ge-1-\mu\frac{\bar{\sigma}}{\sigma}\frac{g_{k+1}^T(g_{k+1}-g_k)}{\|g_{k+1}\|^2}\cdot\frac{g_k^T P_k}{\|g_k\|^2}\ge-1+(-1.2\mu\frac{\bar{\sigma}}{\sigma})(-\sum_{j=0}^{k-1}\bar{\sigma}^j)\\&=-1-1.2\frac{\mu}{\sigma}(-\sum_{j=1}^{k}\bar{\sigma}^j)\ge-1-\sum_{j=1}^{k}\bar{\sigma}^j=-\sum_{j=0}^{k}\bar{\sigma}^j.\end{aligned}$$

从而(8) 式对 $k+1$ 成立. 而

$$-2+\sum_{j=0}^{k}\bar{\sigma}^j<-2+\frac{1}{1-\bar{\sigma}}<0,$$

故(9)式对 $k+1$ 也成立

所以,由归纳法可知,引理1成立

引理1表明,算法A 确定的搜索方向是下降方向

引理2 设点列 $\{x_k\}$ 是由算法A 生成的,在假设(i),(ii)下,若

$$f(x_k)-f(x_{k+1})\geq\delta\left(-\frac{g_k^TP_k}{\|P_k\|}\right),$$

其中 $\delta(\cdot)$ 是强迫函数,则必有

$$\lim_{k\to\infty}\frac{-g_k^TP_k}{\|P_k\|}=0$$

引理3 若 $g(x)$ 在水平集 L_1 上一致连续,则由定义2确定的函数 $S(\cdot)$ 是强迫函数

引理4 在假设(i),(ii)下,对任意的 $x\in L_1$,$P\in R^n$,若 $g^T(x)P<0$,则存在 $\alpha^*>0$,使得

$$f(x+\alpha^*P)=f(x).$$

下面两个定理,将证明算法A 的全局收敛性

定理1 在假设(i),(ii)下,若 α_k 按广义Curry 原则取定,则算法A 确定的点列 $\{x_k\}$ 必满足 $\lim\limits_{k\to\infty}(-g_k^TP_k)/\|P_k\|=0$

证明 由引理1知,对任意 $\bar{\sigma}\in(0,\frac{2}{5})$,有不等式(9)成立 再由引理4知,存在 $\alpha^*>0$ 使

$$f(x_k+\alpha^*P_k)=f(x_k).$$

因此,存在 $\alpha'\in(0,\alpha^*)$ 使得 $g(x_k+\alpha'P_k)^TP_k=0$,从而有不等式

$$g_k^TP_k\leq\mu g_k^TP_k\leq g(x_k+\alpha'P_k)^TP_k=0,\ \mu\in(0,\sigma),$$

由此式可知,存在 $\alpha''\in(0,\alpha')$ 使 $\mu g_k^TP_k=g(x_k+\alpha''P_k)^TP_k$,这表明集合

$$\{\alpha\,|\,g(x_k+\alpha P_k)^TP_k=\mu g(x_k)^TP_k,\alpha>0\}\neq\varnothing.$$

而条件 $|g(x_k+\alpha''P_k)^TP_k|\leq-\sigma g_k^TP_k$ 是显然的 因此广义Curry 原则对所有 k 可以实现

由于 $-g(x_k+\alpha P_k)^TP_k\geq-\mu g_k^TP_k>0,\alpha\in[0,\alpha_k]$,故

$$\varphi'(\alpha)=[f(x_k+\alpha P_k)]'_\alpha=g(x_k+\alpha P_k)^TP_k<0,\tag{11}$$

从而 $\varphi(\alpha)$ 在区间 $[0,\alpha_k]$ 上单调下降 由此推出

$$\varphi(0)-\varphi(\alpha_k)=-\alpha_k\varphi'(\alpha'_k),\quad \alpha'_k\in[0,\alpha_k],$$

即

$$f(x_k)-f(x_{k+1})=-\alpha_kg(\eta_k)^TP_k\geq-\alpha_k\mu g_k^TP_k,\tag{12}$$

其中 η_k 是点 x_k 到 x_{k+1} 联线上的一点 由此推知

$$\begin{aligned}0&<(1-\mu)[-g_k^TP_k]=-g_k^TP_k+\mu g_k^TP_k\\&=-g_k^TP_k+g_{k+1}^TP_k=(g_{k+1}-g_k)^TP_k\leq\|g_{k+1}-g_k\|\cdot\|P_k\|,\end{aligned}$$

从而

$$\|g_{k+1}-g_k\|\geq(1-\mu)\frac{-g_k^TP_k}{\|P_k\|}>0\tag{13}$$

当 $\mu=0$ 时,由 α_k 的取法可知 $g_{k+1}^TP_k=0$ 设 $\bar{\alpha}_k$ 是 $\mu=\frac{\sigma}{2}$ 时由广义Curry 原则确定的步长,记 $\bar{x}_{k+1}=x_k+\bar{\alpha}_kP_k$,则

$$g(\bar{x}_{k+1})^T P_k = \frac{\sigma}{2} g_k^T P_k, \tag{14}$$

其中$\bar{\alpha}_k \in (0, \alpha_k]$, 由(11)式推出: 对任意$\alpha \in [0, \alpha_k]$有

$$f(x_k) \geq f(x_k + \alpha P_k) \geq f(x_{k+1}),$$

因此

$$f(\bar{x}_{k+1}) = f(x_k + \bar{\alpha}_k P_k) \geq f(x_{k+1}).$$

故由(12)式, 有

$$f(x_k) - f(x_{k+1}) \geq f(x_k) - f(\bar{x}_{k+1}) \geq \frac{\sigma}{2}\bar{\alpha}_k(-g_k^T P_k)$$

$$= \frac{\sigma}{2}\bar{\alpha}_k \|P_k\| [(-g_k^T P_k) / \|P_k\|]. \tag{15}$$

但由(13)式及定义2, 有

$$\bar{\alpha}_k \|P_k\| = \|\bar{\alpha}_k P_k\| = \|x_k - x_{k+1}\| \geq S[(1 - \frac{\sigma}{2})(-g_k^T P_k / \|P_k\|)].$$

记 $t = (-g_k^T P_k) / \|P_k\|$, 则由(15)式

$$f(x_k) - f(x_{k+1}) \geq \frac{\sigma}{2} \cdot S((1 - \frac{\sigma}{2}) t) \cdot t \overset{\text{记}}{=} \delta_1(t),$$

则 $\delta_1(t)$为强迫函数, 由引理2可知

$$\lim_{k \to \infty} \frac{-g_k^T P_k}{\|P_k\|} = 0.$$

当 $\mu > 0$时, 由(12)式有

$$f(x_k) - f(x_{k+1}) \geq \alpha_k \mu(-g_k^T P_k) = \mu \|\alpha_k P_k\| [(-g_k^T P_k) / \|P_k\|]$$

而由定义2及(13)式有

$$\|\alpha_k P_k\| = \|x_k - x_{k+1}\| \geq S[(1-\mu)(-g_k^T P_k / \|P_k\|)] = S[(1-\mu)t]$$

从而

$$f(x_k) - f(x_{k+1}) \geq \mu \cdot t \cdot S[(1-\mu)t] \overset{\text{记}}{=} \delta_2(t),$$

则 $\delta_2(t)$是强迫函数, 故由引理2知

$$\lim_{k \to \infty} \frac{-g_k^T P_k}{\|P_k\|} = 0.$$

定理1得证

定理2　在假设(i), (ii)下, $\{x_k\}$是由算法A 生成的点列, 则有

$$\liminf_{k \to \infty} \|g_k\| = 0. \tag{16}$$

证明　用反证法, 若(16)式不成立, 则存在 $\epsilon > 0$使得对所有的 k 有

$$\|g_k\| \geq \epsilon, \tag{17}$$

由于此时定理1的条件也成立, 故有

$$\lim_{k \to \infty} \frac{-g_k^T P_k}{\|P_k\|} = 0. \tag{18}$$

先证明

$$\lim_{k \to \infty} (-g_k^T P_k) = 0. \tag{19}$$

由(18)式，只须证明：对 $k=1,2,\cdots,n$，$\|p_k\|$有界. 事实上，

$$\|p_k\|^2=\|g_k\|^2-2\beta_{k-1}g_k^Tp_{k-1}+\beta_{k-1}^2\|p_{k-1}\|^2$$

$$\leq\|g_k\|^2+2\left|\frac{\beta_{k-1}}{\beta_{k-1}^{PRP}}\right|\bullet\left|\beta_{k-1}^{PRP}\right|\left|g_k^Tp_{k-1}\right|+\left|\frac{\beta_{k-1}}{\beta_{k-1}^{PRP}}\right|^2\bullet(\beta_{k-1}^{PRP})^2\|p_{k-1}\|^2$$

$$\leq\|g_k\|^2+2\bar{\sigma}\frac{\left|g_k^T(g_k-g_{k-1})\right|}{\|g_{k-1}\|^2}(-g_{k-1}^Tp_{k-1})+\left(\frac{\bar{\sigma}}{\sigma}\right)^2\bullet\frac{\left|g_k^T(g_k-g_{k-1})\right|^2}{\|g_{k-1}\|^4}\|p_{k-1}\|^2.$$

由于

$$\left|g_k^T(g_k-g_{k-1})\right|\leq\|g_k\|^2+\|g_k^Tg_{k-1}\|\leq1.2\|g_k\|^2,$$

$$\frac{-g_{k-1}^Tp_{k-1}}{\|g_{k-1}\|^2}\leq\sum_{j=0}^{k-1}\bar{\sigma}^j<\frac{1}{1-\bar{\sigma}}\text{及}\|g_k\|\leq L,$$

故有

$$\|p_k\|^2\leq\|g_k\|^2+\frac{2.4\bar{\sigma}}{1-\bar{\sigma}}\|g_k\|^2+\left(1.2\frac{\bar{\sigma}}{\sigma}\right)^2\frac{\|g_k\|^4}{\|g_{k-1}\|^4}\bullet\|p_{k-1}\|^2$$

$$\leq\frac{1+1.4\bar{\sigma}}{1-\bar{\sigma}}\|g_k\|^2+\left(1.2\frac{\bar{\sigma}}{\sigma}\right)^2\bullet\frac{L^4}{\epsilon^4}\bullet\|p_{k-1}\|^2.$$

记

$$q=\frac{1+1.4\bar{\sigma}}{1-\bar{\sigma}},\quad r=\left(1.2\frac{\bar{\sigma}}{\sigma}\right)^2\bullet\frac{L^4}{\epsilon^4},$$

则有

$$\|p_k\|^2\leq q\|g_k\|^2+r\|p_{k-1}\|^2\leq q\|g_k\|^2+rq\|g_{k-1}\|^2+r^2\|p_{k-2}\|^2$$

$$\cdots\cdots$$

$$\leq q(\|g_k\|^2+r\|g_{k-1}\|^2+r^2\|g_{k-2}\|^2+\cdots+r^{k-1}\|g_1\|^2)$$

$$\leq qL(1+r+r^2+\cdots+r^{k-1})\leq qL(1+r+r^2+\cdots+r^n).$$

这表明：对 $k=1,2,\cdots,n$，$\|p_k\|$有界. 从而有(19)式成立.

现在再证明(16)式成立. 由于

$$\|g_k\|^2=g_k^T(-p_k+\beta_{k-1}p_{k-1})=-g_k^Tp_k+\beta_{k-1}g_k^Tp_{k-1}$$

$$\leq-g_k^Tp_k+\left|\beta_{k-1}\right|(-\sigma g_{k-1}^Tp_{k-1})=-g_k^Tp_k+\left|\frac{\beta_{k-1}}{\beta_{k-1}^{PRP}}\right|\bullet\left|\beta_{k-1}^{PRP}\right|(-\sigma g_{k-1}^Tp_{k-1})$$

$$\leq-g_k^Tp_k+\bar{\sigma}\frac{\left|g_k^T(g_k-g_{k-1})\right|}{\|g_{k-1}\|^2}(-g_{k-1}^Tp_{k-1})\leq-g_k^Tp_k+1.2\bar{\sigma}\bullet\frac{1}{1-\bar{\sigma}}\|g_k\|^2$$

$$=-g_k^Tp_k+\frac{1.2\bar{\sigma}}{1-\bar{\sigma}}\|g_k\|^2,$$

从而对任意 k 有

$$\|g_k\|^2\leq\frac{1-\bar{\sigma}}{1-2.2\bar{\sigma}}(-g_k^Tp_k),$$

故有

$$\lim_{k\to\infty}\|g_k\|^2\leq\frac{1-\bar{\sigma}}{1-2.2\bar{\sigma}}\lim_{k\to\infty}(-g_k^Tp_k)=0.$$

这与假设(17)矛盾. 所以(16)式成立，即

$$\liminf_{k\to\infty}\|g_k\| = 0$$

定理2得证

参 考 文 献

1 Fletcher R, Reeves C M. *Function minimization by conjugate gradients*. Computer Journal, 1964, **7**: 149-154

2 Polak E, Ribierè G. *Note sur la convergences de méthods des directions conjugés*. Rev. Fr. Inr. Rech. Oper., 1969, **16**: 35-43

3 Powell M J D. *Nonconvex minimization calculation and the conjugate gradient method*. In Numerical Analysis, Dundee, 1983, Griffiths D. F ed.

4 Al-Baali M. *Descent Property and global convergence of the Fletcher-Reeves Method with inexact line searches*. IMA Journal of Numerical Analysis, 1985, **5**(1): 121-124

5 袁亚湘 非线性规划的数值方法 上海科学技术出版社, 1993

6 Powell M J D. *Nonconvex minimization calculations and the conjugate gradient method*. Numerical Analysis, Lecture Notes in Mathematics, 1066, Springer-Verlag, Berlin, (D. F. Griffiths, ed.), 1984, 122-141

7 Grippo L and Lucidi S. *A global convergence properties of conjugate gradient methods for optimization*. SIAM, 1995, **71**: 399-405

8 席少霖 非线性最优化方法 北京: 高等教育出版社, 1992, 169-183

9 赵瑞安、吴方. 非线性最优化理论和方法 杭州: 浙江省科学技术出版社, 1992, 21-26

Global Convergence of the Restarting Conjugate Gradient Algorithm with a Generalized Curry Linesearch

Jiao Baocong *Chen Lanping*

(Dept. of Math., Capital Normal University, 100037)

Abstract

This paper presents a new restarting conjugate gradient algorithm for unconstrained optimization problem: $\min f(x)$, $x \in R^n$. This algorithm used a generalized Curry linesearch, parameter β_k can be selected in a finite closed interval. Especially, it is allowed that β_k is negative. The global convergence of this algorithms is proved under weaker conditions.

Keywords restarting conjugate gradient algorithm, generalized Curry linesearch, global convergence