DOI:10.13656/j.cnki.gxkx.2013.01.014

广西科学 Guangxi Sciences 2013,20(1):5~8

一种修改的 PRP 共轭梯度法*

A Modification of PRP Conjugate Gradient Method

黎 勇

LI Yong

(百色学院数学与计算机科学系,广西百色 533000)

(Department of Mathematics and Computer Science, Baise University, Baise, Guangxi, 533000, China)

摘要:用新的 PRP 参数公式修改一种已知的线搜索,建立此线搜索下的共轭梯度算法,并证明算法能满足充分下降条件,而且在适当条件下全局收敛.

关键词:无约束优化 共轭梯度法 非精确线搜索 全局收敛性

中图法分类号:O224 **文献标识码**:A **文章编号**:1005-9164(2013)01-0005-04

Abstract: A conjugate gradient method is presented with a known line search modified by a new PRP parameter formula, and it is proved that the sufficient descent condition is held and the global convergence is established under some proper conditions.

Key words: unconstrained optimization, conjugate gradient method, inexact line search, global convergence

对无约束优化问题 $\min\{f(x) \mid x \in \Re^n\}$,用非线性共轭梯度法求解比较常用而且有效.该方法的迭代公式为

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

$$d_k = \begin{cases} -g_k, k=1, \\ -g_k + \beta_k d_{k-1}, k \geqslant 2. \end{cases} \tag{2}$$

其中 t_k 是步长,d_k 是搜索方向,$g_k = \nabla f(x_k)$ 是 $f(x)$ 在点 x_k 处的梯度,β_k 是参数,不同的参数选取对应着不同的共轭梯度法.已报道的著名的迭代参数公式[1~4]有 $\beta_k^{\mathrm{HS}} = \dfrac{g_k^T(g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}$,$\beta_k^{\mathrm{FR}} = \dfrac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$,$\beta_k^{\mathrm{PRP}} = \dfrac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}$.

对线搜索型的共轭梯度法来说,步长因子 t_k 的计算很重要,即从 x_k 沿 d_k 方向如何寻找一个"好"的点作为下一个迭代点.目前较常用的线搜索条件[5]有:弱 Wolfe-Powell 型线搜索,强 Wolfe-Powell 型线搜索,Armijo 型线搜索,Armijo-Goldstein 型线搜索等.近期,文献[6]提出了一种新的的线搜索 Armijo-type line search(ATLS),具体描述如下:

设 $f_k = f + \dfrac{1}{2}(x - x_k)^T B_k (x - x_k)$,其中 B_k 是对称正定矩阵.令 $\alpha \in [0, \dfrac{1}{2})$,$c \in (0,1)$,$\mu > 0$.选取 $t_k = \rho^{j_k}$,其中 j_k 是使得下式成立的最小非负整数 j:

$$f(x_k + \rho^j d_k) - f(x_k) \leqslant \alpha\rho^j g_k^T d_k - \frac{\mu}{2}(\rho^j)^2 \| d_k \|^2,$$

且 $g(x_k + \rho^j d_k)^T Q_k^{\mathrm{PRP}}(j) \leqslant -c \| g(x_k + \rho^j d_k) \|^2$,其中 $Q_k^{\mathrm{PRP}}(j)$ 为

$$Q_k^{\mathrm{PRP}}(j) = -g(x_k + \rho^j d_k) + \frac{g(x_k + \rho^j d_k)^T (g(x_k + \rho^j d_k) - g_k)}{\| g_k \|^2} d_k\ .$$

他们证明 PRP 方法在该线搜索下满足充分下降条件:

$$g_k^T d_k \leqslant -c \| g_k \|^2, c \in (0,1)\ , \tag{3}$$

而且在此基础上,算法的全局收敛容易被证明.

受文献[6]的启发,本文结合文献[7]中提出的新的修正 PRP 参数公式:

$$\beta_k^* = \frac{g_k^T\left(\dfrac{\| g_{k-1} \|}{\| g_k \|} g_k - g_{k-1}\right)}{g_{k-1}^T g_{k-1}}\ , \tag{4}$$

重新定义 $Q_k^*(j)$,从而得到一种新的 ATLS 线搜索

收稿日期:2012-03-27

修回日期:2012-04-23

作者简介:黎 勇(1973-),男,副教授,硕士,主要从事最优化理论的研究。

*国家自然科学基金项目(No. 10761001);广西教育厅科研项目(No. 201010LX501);广西高校优秀人才计划项目(201261)资助。

($*$ATLS),并建立了此线搜索下的共轭梯度算法.

1 新的线搜索及算法

令 $t_k=\rho^{j_k}$,其中 j_k 是使得下式成立的最小非负整数 j:

$$f(x_k+\rho^j d_k)-f(x_k)\leqslant\alpha\rho^j g_k^T d_k-\frac{\mu}{2}(\rho^j)^2\|d_k\|^2,\tag{5}$$

而且

$$g(x_k+\rho^j d_k)^T Q_k^*(j)\leq -c\|g(x_k+\rho^j d_k)\|^2,\tag{6}$$

其中 $Q_k^*(j)$ 定义为

$$Q_k^*(j)=-g(x_k+\rho^j d_k)+\frac{g(x_k+\rho^j d_k)^T\left(\frac{\|g_k\|}{\|g(x_k+\rho^j d_k)\|}g(x_k+\rho^j d_k)-g_k\right)}{\|g_k\|^2}d_k.\tag{7}$$

再假定目标函数 $f(x)$ 满足如下基本假设:

假设 1 水平集 $\Omega=\{x\in\Re^n:f(x)\leqslant f(x_1)\}$ 有下界.

假设 2 $f(x)$ 在 Ω 的一个邻域 N 内连续可微,其梯度 Lipschitz 连续.即存在一个常数 $L>0$,使得对 $\forall x,y\in\Omega$,有 $\|g(x)-g(y)\|\leqslant L\|x-y\|$.

引理 1 如果对 $\forall k\in N$,$g_k^T d_k<0$,则存在非负整数 j_k,使得 $t_k=\rho^{j_k}$ 满足$*$ATLS 线搜索(5)~(7).

引理 1 的证明与文献[6]中 Lemma 2.1 的证明类似,在此省略具体证明过程.

算法 1

步骤 1 给定初值 $x_1\in\Re^n$,$\alpha\in[0,\frac{1}{2})$,$c\in(0,1)$,$\mu>0$.令 $d_1=-g_1$,$k=1$.若 $g_1=0$,则停止,否则转步骤 2.

步骤 2 求出满足$*$ATLS 线搜索(5)~(7)的步长 $t_k>0$.

步骤 3 迭代运算 $x_{k+1}=x_k+t_k d_k$,$g_{k+1}=g(x_{k+1})$.若 $g_{k+1}=0$,则停止,否则转步骤 4.

步骤 4 由(4)式计算 β_k,由(2)式计算 d_{k+1}.

步骤 5 令 $k:=k+1$,转步骤 3.

2 算法的全局收敛性

引理 2 考虑算法 1,对 $\forall k\in N$,若 $g_k^T d_k<0$,则

$$g_{k+1}^T d_{k+1}\leqslant -c\|g_{k+1}\|^2.\tag{8}$$

证明 根据引理 1,通过$*$ATLS 线搜索产生步长 t_k,并由算法 1 得到相应的 x_{k+1},g_{k+1},β_k 和 d_{k+1},因此

$$g_{k+1}^T d_{k+1}=-\|g_{k+1}\|^2+\beta_{k+1}^* g_{k+1}^T d_k=$$

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

$$-\|g(x_k+t_k d_k)\|^2+\frac{g(x_k+t_k d_k)^T\left(\frac{\|g_k\|}{\|g(x_k+t_k d_k)\|}g(x_k+t_k d_k)-g_k\right)}{\|g_k\|^2}\cdot g(x_k+x_k+t_k d_k)^T d_k,$$

因为 t_k 满足(6)式,所以

$$-\|g(x_k+t_k d_k)\|^2+\frac{g(x_k+t_k d_k)^T\left(\frac{\|g_k\|}{\|g(x_k+t_k d_k)\|}g(x_k+t_k d_k)-g_k\right)}{\|g_k\|^2}\cdot g(x_k+t_k d_k)^T d_k\leqslant -c\|g(x_k+t_k d_k)\|^2.$$

即(8)式成立.

定理 1 对算法 1,如果 $g_k\neq 0$,则对 $\forall k>0$,充分下降条件(3)式成立.

证明 假设 $g_1\neq 0$,因为 $d_1=-g_1$,所以 $g_1^T d_1=-\|g_1\|^2\leqslant -c\|g_1\|^2$.若 $g_k\neq 0$,则 $g_k^T d_k\leqslant -c\|g_k\|^2<0$,其中 $c\in(0,1)$.由引理 2 知,$g_{k+1}^T d_{k+1}\leqslant -c\|g_{k+1}\|^2$.再利用数学归纳法容易得出定理 1.

引理 3 如果假设 1 成立,则有

$$\lim_{k\to\infty}t_k\|d_k\|=0,\tag{9}$$

$$\lim_{k\to\infty}-t_k g_k^T d_k=0.\tag{10}$$

证明 由假设 1 有

$$\sum_{k=1}^{\infty}(f(x_k)-f(x_{k+1}))=\lim_{N\to\infty}\sum_{k=1}^{N}(f(x_k)-f(x_{k+1}))=\lim_{N\to\infty}(f(x_1)-f(x_{k+1}))=f(x_1)-f_1,$$

所以 $\sum_{k=1}^{\infty}(f(x_k)-f(x_{k+1}))<+\infty$.又由(5)式知

$$f(x_k+t_k d_k)-f(x_k)\leqslant\alpha t_k g_k^T d_k-\frac{\mu}{2}(t_k)^2\|d_k\|^2,$$

所以 $\sum_{k=1}^{\infty}t_k^2\|d_k\|^2<\infty$,$\sum_{k=1}^{\infty}-t_k g_k^T d_k<\infty$.

这样容易得到(9)式和(10)式成立.

引理 4 如果假设 1 和 2 都成立,序列 $\{x_k\}$ 由算法 1 产生.则存在常数 $M_1>0$ 使得对 $\forall k$,

$$t_k\geqslant M_1\|g_k\|^2/\|d_k\|^2.$$

引理 4 的证明可以参考文献[6]中 Lemma3.3.

引理 5 如果假设 1 和 2 都成立,且对 $\forall k$,$\exists\varepsilon>0$,使得

$$\|g_k\|\geqslant\varepsilon.\tag{11}$$

则存在常数 $M_2 > 0$，使得对 $\forall k$，

$$\| d_k \| \leqslant M_2 . \tag{12}$$

证明 利用 Cauchy-schwartz 不等式，由(2)式有

$$\| d_k \| \leqslant \| g_k \| + | \beta_k^* | \| d_{k-1} \| \leqslant \| g_k \| + \frac{g_k^T (\frac{\| g_{k-1} \|}{\| g_k \|} g_k - g_{k-1})}{\| g_{k-1} \|^2} \| d_{k-1} \| \leqslant \| g_k \| + \frac{\| g_k \| \cdot \| \frac{\| g_{k-1} \|}{\| g_k \|} g_k - g_k + g_k - g_{k-1} \|}{\| g_{k-1} \|^2} \cdot \| d_{k-1} \| \leqslant \| g_k \| + \frac{\| g_k \| (\| \frac{\| g_{k-1} \|}{\| g_k \|} g_k - g_k \| + \| g_k - g_{k-1} \|)}{\| g_{k-1} \|^2} \cdot \| d_{k-1} \| \leqslant \| g_k \| + \frac{2 \| g_k \| \| g_k - g_{k-1} \|}{\| g_{k-1} \|} \| d_{k-1} \| .$$

再利用假设 2 和(11)式，可得

$$\| d_k \| \leqslant \| g_k \| + \frac{2L \| g_k \| \| x_k - x_{k-1} \|}{\| g_{k-1} \|} \| d_{k-1} \| \leqslant \| g_k \| + \| g_k \| \frac{2L t_{k-1} \| d_{k-1} \|}{\varepsilon} \| d_{k-1} \| . \tag{13}$$

另外，因为序列 $\{x_k\}$ 有界，又根据假设 2，可以推出：存在 $M_3 > 0$，使得对 $\forall k$，

$$\| g_k \| \leqslant M_3 . \tag{14}$$

所以根据(13)式和(14)式得

$$\| d_k \| \leqslant M_3 + \frac{2M_3 L}{\varepsilon} t_{k-1} \| d_{k-1} \|^2 = M_3 + (\frac{2M_3 L}{\varepsilon} t_{k-1} \| d_{k-1} \|) \| d_{k-1} \| . \tag{15}$$

而由引理 3 的(9)式容易推知：存在一个常数 $q \in (0, 1)$ 和一个整数 k_0，使得对 $\forall k \geqslant k_0$，有 $\frac{2M_3 L}{\varepsilon} t_{k-1} \| d_{k-1} \| \leqslant q$. 所以对 $\forall k > k_0$，由(15)式有

$$\| d_k \| \leqslant M_3 + q \| d_{k-1} \| \leqslant M_3 (1 + q + q^2 + \cdots + q^{k-k_0-1}) + q^{k-k_0} \| d_{k_0} \| \leqslant \frac{M_3}{1-q} + q^{k-k_0} \| d_{k_0} \| \leqslant \frac{M_3}{1-q} + \| d_{k_0} \| .$$

令 $M_2 = \max\{ \| d_1 \|, \| d_2 \|, \cdots, \| d_{k_0} \|, \frac{M_3}{1-q} + \| d_{k_0} \| \}$，则(12)式对 $\forall k$ 成立.

定理 2 如果假设 1 和 2 都成立，序列 $\{x_k\}$ 由算法 1 产生，则

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

证明 反证法. 假设结论不成立，则存在常数 $\varepsilon > 0$，使得对 $\forall k$，

$$\| g_k \| \geqslant \varepsilon . \tag{17}$$

由引理 4 知，存在常数 $M_1 > 0$ 使得对 $\forall k, t_k \geqslant M_1 \| g_k \|^2 / \| d_k \|^2$，再结合引理 5，可得 $\| g_k \|^2 \leqslant \frac{M_2}{M_1} t_k \| d_k \|$. 令 $k \to \infty$，利用(9)式，则从以上不等式可以推出与(17)式相矛盾的结果. 故(16)式成立.

3 数值实验

实验环境为 Windows XP + Matlab6.5，CPU 2.8G，RAM 448M，算法线搜索参数为 $\alpha = 10^{-3}$，$\rho = 0.5$，$c = 0.3$，$\mu = 0.1$，$\varepsilon = 10^{-5}$，内循环 Armijo 搜索次数限制在 10^3 以内；终止条件为 $\| g_k \| \leqslant \varepsilon$，或总迭代次数 $N_i > 2 \times 10^4$，或函数计算次数 $Nf > 10^5$. 测试结果见表 1. 表 1 中其它符号表示如下：

Problem 表示测试函数的名称，Dim 为相应的维数；Ng 为目标梯度计算次数 ($Ng = Nf$)；time 为 CPU 时间(秒)；f^* 为算法满足精度要求 $\| g_k \| \leqslant \varepsilon$ 而终止时，所得近似最优点的函数值.

表 1 数值实验结果

Table 1 Numberical result

Problem	Dim	$Ni/Nf(Ng)$/time	f^*
ROSE	2	90/813/0.2190	5.793663e−011
FROTH	2	92/991/0.2660	6.656854e−013
BEALE	2	35/185/0.0620	6.889297e−012
JENSAM	2	58/433/0.1410	2.653333e−001
HELIX	3	1358/12713/4.0620	1.999463e−011
BARD	3	1044/4669/1.8440	8.214880e−003
GAUSS	3	6/21/0.0160	1.147093e−008
GULF	3	2/9/0.0780	3.850000e−002
SING	4	1931/13934/3.8440	2.915261e−008
WOOD	4	177/1632/0.4530	1.670416e−011
KOWOSB	4	3140/9199/3.4680	3.075168e−004
OSB2	11	1408/7561/9.0630	4.013774e−002
ROSEX	8	141/1277/0.3590	1.549463e−011
ROSEX	50	147/1334/0.5160	2.219352e−011
ROSEX	100	126/1139/0.6400	1.139490e−011
SINGX	8	2815/20243/6.0790	3.177070e−008
PEN1	2	7/16/0.0150	3.591560e−005
PEN2	4	32/146/0.0630	1.037527e−005
PEN2	50	1095/9102/9.0470	4.296098e+000
VARDIM	2	12/68/0.0310	4.841895e−013
VARDIM	50	21/430/0.1720	4.154630e−016
TRIG	3	38/55/0.0310	2.573685e−003
TRIG	50	86/92/0.2190	5.452389e−006

续表 1

Continue table 1

Problem	Dim	$Ni/Nf(Ng)$ /time	f^*
TRIG	100	93/100/0.9220	2.406033e−006
BV	3	27/125/0.0460	6.526502e−012
BV	10	266/1173/0.4070	3.274220e−010
IE	3	11/34/0.0150	6.738058e−012
IE	50	13/40/0.1100	1.055150e−011
IE	100	13/40/0.3430	2.097808e−011
IE	200	14/43/1.3750	8.520314e−012
IE	500	14/43/8.7820	2.124331e−011
TRID	3	29/144/0.0470	9.229866e−001
TRID	50	85/579/0.2500	1.407812e+000
TRID	100	79/546/0.5000	1.407812e+000
TRID	200	85/581/0.6870	1.407812e+000
BAND	3	10/76/0.0310	2.157846e−013
BAND	50	39/316/0.2820	4.097346e−013
LIN	2	1/3/0.0160	1
LIN	50	5/15/0.0470	5.000000e+001
LIN	500	17/51/5.5930	5.000000e+002
LIN	1000	1/3/0.6410	1.000000e+000
LIN1	2	21/190/0.0620	4.285714e−001
LIN1	10	19/381/0.1250	2.391304e+000

从表 1 的结果看出，对上述测试函数而言，本文提出的新算法具有比较好的计算效果，适用于大规模科学计算.

致谢：

衷心感谢审稿专家提出的修改意见！

参考文献：

[1] Hestenes M R，Stiefel E L. Methods of conjugate gradients for solving linear systems[J]. J Res Nat Bur Standards Sect，1952，5(49)：409-436.

[2] Fletcher R，Reeves C. Function minimization by conjugate gradients[J]. ComPuter Journal，1964，7(2)：149-154.

[3] Polak E，Ribière G. Note Sur la convergence de directions conjugèes[J]. Rev Francaise informat Recherche Operationelle，3e Annèe，1969，16：35-43.

[4] Polyak B T. The conjugate gradient method in extreme problem[J]. USSR Comp Math and Math Phys，1969，9(4)：94-112.

[5] 戴彧虹，袁亚湘. 非线性共轭梯度法[M]. 上海：上海科技出版社，1999.

[6] Wei Z，Li G，Qi L. Global convergence of the Polka-Ribière-Polyak conjugate gradient method with an Armijo-type inexact line search for nonconvex unconstrained optimization problem[J]. Mathematics of Computation，2008，77(6)：2173-2193.

[7] 黎勇. 一类修正 PRP 共轭梯度法的全局收敛性及其数值试验结果[J]. 西南大学学报：自然科学版，2011，33(11)：23-28.

（责任编辑：尹　闯）

美国研制出廉价石墨烯海绵传感器

最近几年，在操作纳米结构并用其制造性能卓越的探测器以精确追踪空气中的化学物质方面，科学家们已经取得了重大的进步。他们已经研制出各式各样的传感器。然而，传感器的设计都非常复杂，常常依赖单个纳米结构，而且，科学家们需要对这样的结构进行仔细操作以及更加精确的分析，制造出的传感器还往往不能重复使用，而且必须在特定的温度或压力下才能工作，因此，科学家们一直没有制造出一款可靠、便宜而且可以重复使用的手持传感设备。现在，英国科学家们使用石墨烯泡沫研制出一种邮票大小的新型传感器。科学家们将石墨烯，即单层碳原子，种植在泡沫镍结构上，随后移除泡沫镍，留下一个类似泡沫的石墨烯结构，其具有独特的电性，能够用于执行传感任务。将新型传感器暴露于空气中时，空气中的粒子会被吸收到泡沫表面，而且每个这样的粒子会用不同的方式影响石墨烯泡沫，对其电阻进行微小改动。让电流通过其中并测量电阻变化，就能知道泡沫上依附的是什么粒子。科学家们让大约 100mA 电流通过该泡沫，发现石墨烯泡沫能够导致粒子解吸，也就是说，粒子自动从传感器上剥落下来，清除这些粒子，传感器就可以重复使用了。

石墨烯泡沫非常容易处理，操作简单，在室温下也能很好地工作。这都是科学家们非常心仪的特质。该石墨烯泡沫传感器可以让科学家们更快制造出更便宜实用的手持传感设备来对大气进行探测。

（据科学网）