

具有充分下降性的修正 PRP 算法及其收敛性^{*}

喻高航, 关履泰

(中山大学科学计算与计算机应用系, 广东 广州 510275)

摘要: 共轭梯度法因其算法简单、存储需求小, 非常适合于求解大规模优化问题。在所有的共轭梯度法中, PRP 方法被认为是数值表现最好的方法之一。然而, 对一般非凸函数, PRP 方法即使采用精确线搜索也不能保证全局收敛。本文基于一个修正的 PRP 公式, 提出了一类无需线搜索而具有充分下降性的共轭梯度算法。在一定条件下, 建立了该算法的全局收敛性结果。数值试验表明这种改进是有效的。

关键词: 无约束优化; 大规模优化; 共轭梯度方法; 全局收敛性

中图分类号: O224 O241.5 **文献标识码:** A **文章编号:** 0529-6579 (2006) 04-0011-05

1 研究背景

共轭梯度法具有算法简便, 存储需求小等优点, 十分适合于求解大规模优化问题。考虑无约束优化问题

$$\min \{ f(x) \mid x \in R^n \} \quad (1)$$

这里 $f(x)$ 连续可微。求解问题 (1) 的共轭梯度法具有如下形式:

$$x_{k+1} = x_k + \alpha_k d_k \quad (2)$$

和

$$d_k = \begin{cases} -g_k & \text{if } k=1 \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 2 \end{cases} \quad (3)$$

其中, $\alpha_k > 0$ 是步长因子; $g_k = \nabla f(x_k)$; β_k 为一参数。以下是一个著名的选取 β_k 的共轭梯度公式

$$\beta_k^{\text{PRP}} = \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}} \quad (4)$$

由 Polak-Ribière^[1] 和 Polyak^[2] 于 1969 提出, 故相应的算法称为 PRP 算法。

共轭梯度法在不同类型的线搜索下的收敛性已被广泛研究^[3-6]。在所有的共轭梯度法中, PRP 方法被认为是数值表现最好的方法之一。但对一般非凸函数, Powell^[7] 举出了反例表明即使采用精确线搜索 PRP 方法也不具有全局收敛性。如果使用非精确线搜索, Dai^[8] 举出了例子表明, 即使目标函数 $f(x)$ 一致凸, 而且强 Wolfe 线搜索中的参数 $\sigma \in (0, 1)$ 充分小, PRP 方法都有可能产生一个上升的搜索方向。对一般非凸函数, Powell^[9] 建议限制 PRP 方法中参数 β_k 为非负: $\beta_k^{\text{PRP}} = \max\{0, \beta_k^{\text{PRP}}\}$ 。

Gilbert 和 Nocedal^[4] 考虑了 Powell 的上述建议, 并在适当的线搜索条件下, 建立了上述修正 PRP 方法对一般非凸函数的全局收敛结果。

为保证算法的全局收敛性和有效性, 在共轭梯度算法的实现过程中步长 α_k 通常由 Wolfe 线搜索条件所决定, 即确定步长 $\alpha_k > 0$ 满足

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \quad (5)$$

以及

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \quad (6)$$

其中, $\delta \in (0, 1)$, $\sigma \in (\delta, 1)$ 。此外, 共轭梯度法的收敛性分析中经常要用到充分下降条件

$$g_k^T d_k \leq -\varsigma \|g_k\|^2 \quad (7)$$

其中, $\varsigma > 0$ 为某常数。比如, Al-Baali^[3], Gilbert 和 Nocedal^[4] 等, 在分析强 Wolfe 线搜索下共轭梯度算法的全局收敛性时, 充分下降条件 (7) 都起到关键的作用。为了满足条件 (7), 现有的主要方法大都依赖于具体的非精确线搜索技术^[4-6], 有时导致线搜索非常复杂。本文主要讨论了一类不依赖于具体的线搜索能保证充分下降的修正 PRP 共轭梯度算法及其全局收敛性。最后, 对一些典型的数值算例进行实验, 结果表明是有效的。

2 算法及其全局收敛性分析

考虑共轭梯度算法 (2)–(3), 其中 $\beta_k = \beta_k^{\text{DIRP}}$ 由如下公式迭代 (Dirp 为 descend 的第一个字母):

$$\beta_k^{\text{DIRP}} = \beta_k^{\text{PRP}} - \frac{\|y_{k-1}\|^2}{\|g_{k-1}\|^4} g_k^T d_{k-1} \quad (8)$$

* 收稿日期: 2005-07-04

基金项目: 国家自然科学基金资助项目 (10572154); 香港中山大学高等学术研究中心资助项目 (05M7)

作者简介: 喻高航 (1975 年生), 男, 博士生; 通讯联系人: 关履泰; E-mail: maghyu@163.com

这里 $y_{k-1} = g_k - g_{k-1}$, $C > 1/4$. 将该算法简称 DPRP 方法. 不需要任何线搜索条件, (8) 式能保证 DPRP 方法的充分下降性, 见如下定理.

定理 1 考虑方法 (2) - (3), 其中 $\beta_k = \beta_k^{\text{DPRP}}$ 由 (8) 生成. 如果 $\|g_{k-1}\| \neq 0$, 则对所有的 $k \geq 1$ 都有

$$g_k^T d_k \leq -\left(1 - \frac{1}{4C}\right) \|g_k\|^2 \quad (9)$$

证明 当 $k=1$ 时有 $g_1^T d_1 = -\|g_1\|^2$, 显然 (9) 成立. 当 $k > 1$ 时, 由 (3), (8) 有,

$$g_k^T d_k = -\|g_k\|^2 +$$

$$g_k^T d_{k-1} \left(\frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} - \frac{C\|y_{k-1}\|^2}{\|g_{k-1}\|^4} g_k^T d_{k-1} \right) =$$

$$\frac{g_k^T d_{k-1} g_k^T y_{k-1} - \|g_k\|^2 \|g_{k-1}\|^2 - \frac{C\|y_{k-1}\|^2}{\|g_{k-1}\|^2} (g_k^T d_{k-1})^2}{\|g_{k-1}\|^2}$$

设 $u = \frac{\|g_{k-1}\|}{\sqrt{2C}} g_k$ 以及 $v = \frac{\sqrt{2C} g_k^T d_{k-1}}{\|g_{k-1}\|} y_{k-1}$, 则有

$$g_k^T d_k = \frac{u^T v - \frac{1}{2} (\|u\|^2 + \|v\|^2)}{\|g_{k-1}\|^2} + \frac{-(1 - \frac{1}{4C}) \|g_k\|^2 \|g_{k-1}\|^2}{\|g_{k-1}\|^2}$$

结合不等式 $u^T v \leq \frac{1}{2} (\|u\|^2 + \|v\|^2)$, 知不等式 (9) 对所有的 $k \geq 1$ 都成立.

为讨论 DPRP 方法的收敛性, 假定目标函数 $f(x)$ 满足如下基本假设:

假设 1

① 水平集 $\Omega = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_1)\}$ 有下界; ② $f(x)$ 在 Ω 的一个邻域 N 内连续可微, 且其导数 L -pschitz 连续, 存在常数 $L > 0$ 使得

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in N \quad (10)$$

在假设 1 的前提下, 可以给出 DPRP 方法的一个收敛定理如下.

定理 2 若假设 1 成立, 且步长 α_k 满足 (5), 以及对所有的 $k \geq 1$ 都存在常数 $\zeta > 0$, 使 $\alpha_k \geq \zeta$ 成立, 则对 DPRP 算法产生的点列 $\{x_k\}$ 有: 或者存在某个 k 有 $\|g_k\| = 0$, 或者

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \quad (11)$$

证明 由 (5) 和 (9) 可知, 对所有的 $k \geq 1$ 都有 $x_k \in \Omega$. 根据水平集 Ω 有下界和 $f(x)$ 的连续性知序列 $\{f(x_k)\}$ 的极限存在. 由关系式 (5) 易证

$$\sum_{k=1}^{\infty} \|g_k\| \leq \sum_{k=1}^{\infty} -(f(x_{k+1}) - f(x_k)) < \infty \quad (12)$$

进一步, 由 (9) 和 $\alpha_k \geq \zeta$ 可知 (11) 成立. 证毕.

文 [10] 中, Dai 给出如下一个 Amijo 类型的线搜索 (简记为 ALS 线搜索): 给定常数 $\lambda \in (0, 1)$ 以及 $\zeta \in (0, 1]$, 确定最小的整数 $m \geq 0$ 使得 $\alpha_k = \lambda^m$ 满足 (5) 和

$$g(x_k + \alpha_k d_k)^T d(x_k + \alpha_k d_k) \leq -\zeta \|d(x_k + \alpha_k d_k)\|^2 \quad (13)$$

文 [10] 证明了 PRP 方法在上述 Amijo 类型的线搜索下具有 $\lim_{k \rightarrow \infty} \|g_k\| = 0$ 意义上的收敛性. 下面, 证明对 DPRP 方法也有相应的结果.

引理 1 设目标函数满足假设 1, 考虑 DPRP 方法, 其中步长 α_k 由 ALS 线搜索所得, 则这样的 α_k 一定存在且对所有的 $k \geq 1$, 存在常数 $\zeta > 0$ 使 $\alpha_k \geq \zeta$ 成立.

证明 与文 [10] 中的引理 2.8 类似. 由 $d = -g$, 知

$$g_k^T d_k \leq -\zeta \|d_k\|^2 \quad (14)$$

对 $k=1$ 时成立. 现假设 (14) 对某个 k 成立, 则有 $\|d_k\| \leq \zeta^{-1} \|g_k\|$. 对任何 $\alpha > 0$, 定义 $x_{k+1} = x_k + \alpha d_k$, $g_{k+1} = g(x_{k+1})$ 以及 $\tilde{d}_{k+1} = -g_{k+1} + \beta_{k+1}^{\text{DPRP}} d_k$. 根据中值定理和 (10), 有

$$f(x_{k+1}) - f = \int_0^1 g(x_k + t\alpha d_k)^T (\alpha d_k) dt =$$

$$\alpha g_k^T d_k + \int_0^1 [g(x_k + t\alpha d_k) - g_k]^T (\alpha d_k) dt \leq$$

$$\alpha g_k^T d_k + \frac{1}{2} L \alpha^2 \|d_k\|^2$$

因此可得 $f(x_{k+1}) - f \leq \delta \alpha g_k^T d_k$ 对所有的 $\alpha \in$

$(0, \frac{2(1-\delta)}{L} \frac{|g_k^T d_k|}{\|d_k\|^2}]$ 成立. 由假设 (14) 可得

$$f(x_{k+1}) - f \leq \delta \alpha g_k^T d_k$$

$$\alpha \in (0, \frac{2\zeta(1-\delta)}{L}] \quad (15)$$

另一方面, 由 $\tilde{d}_{k+1}$ 的定义和 (10) 可得

$$\|\tilde{d}_{k+1}\| \leq \|g_{k+1}\| \left[1 + \frac{L \alpha \|d_k\|^2}{\|g_k\|^2} + \frac{C L^2 \alpha^2 \|d_k\|^4}{\|g_k\|^4} \right] \quad (16)$$

和

$$\begin{aligned} g_{k+1}^T \tilde{d}_{k+1} &= -\|g_{k+1}\|^2 \left[1 - \frac{g_{k+1}^T (g_{k+1} - g_k)}{\|g_k\|^2} \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2} + \right. \\ &\quad \left. \frac{C\|g_{k+1} - g_k\|^2}{\|g_k\|^4} \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2} \right] \leq \\ &= -\|g_{k+1}\|^2 \left[1 - \frac{L \alpha \|d_k\|^2}{\|g_k\|^2} \right] \end{aligned}$$

因此，选取某个常数 $\xi > 0$ 使得 $1 - \xi \geq \xi(1 + \xi + C_4^2)$ ，由 (16) 和上式，有

$$\begin{aligned} \xi_{k+1}^T \tilde{d}_{k+1} &\leq -\xi \|\tilde{d}_{k+1}\|^2, \\ \alpha &\in (0, \frac{\xi \|\tilde{g}_k\|^2}{\|\tilde{d}_k\|^2}] \end{aligned} \tag{17}$$

根据 $\|\tilde{d}_k\| \leq \xi^{-1} \|\tilde{g}_k\|$ ，(15) 和 (17)，令 $\xi = \min\{1, \frac{2\lambda \xi(1-\delta)}{L}, \frac{\lambda \xi^2}{L}\}$ ，知结论成立，证毕。

显然，由引理 1 和定理 2 可知如下命题成立。
定理 3 设目标函数满足假设 1，考虑 DPRP 方法，其中步长 α_k 由 ALS 线搜索所得，则对算法产生的点列 $\{\tilde{x}_k\}$ 有：或者存在某个 k 有 $\|\tilde{g}_k\| = 0$ ，或者 (11) 成立。

如上述，在一定的条件下，已经建立了 DPRP 方法的全局收敛性。但是 ALS 线搜索有一个明显的缺点，它不能保证 (较为精确的线搜索，也就是说它会拒绝一个已得的好的试探步长 (比如局部极小点或满足 Wolfe 线搜索的点)。另外，对共轭梯度法，限制步长 α_k 总是小于或等于 1，会制约算法的效率。为克服上述缺点，结合 Wolfe 线搜索技术，给出如下算法：

DPRP-WLS 算法

给定数据: $\alpha_{\min}, \xi, \lambda, \delta \in (0, 1), \sigma \in (\delta, 1), C > 1/4, \epsilon > 0$ 以及 $\tilde{x} \in R^n$.

初始化. 令 $\tilde{d} = -\tilde{g}$ 以及 $k \leftarrow 1$.

While $\|\tilde{g}_k\| \geq \epsilon$ do

Step1. 计算 $\alpha_k > 0$ 使其满足 Wolfe 线搜索条件 (5) 和 (6);

Step2

If $\alpha_k \geq \alpha_{\min}$ then

令 $\tilde{x}_{k+1} = \tilde{x}_k + \alpha_k \tilde{d}_k; \tilde{d}_{k+1} = -\tilde{g}_{k+1} + \beta_{k+1}^{\text{DPRP}} \tilde{d}_k$.

Else 由 ALS 线搜索生成步长: $\alpha_k = \lambda^m$ 满

足 (5) 和 (13); 令 $\tilde{x}_{k+1} = \tilde{x}_k + \alpha_k \tilde{d}_k$; 以及

$$\tilde{d}_{k+1} = -\tilde{g}_{k+1} + \beta_{k+1}^{\text{DPRP}} \tilde{d}_k$$

End if

Step3 Set $k = k + 1$.

由 DPRP-WLS 算法的 Step 2 知，对所有的 $k \geq 1$ 都存在常数 $\xi = \min\{\alpha_{\min}, \xi\}$ ，使 $\alpha_k \geq \xi$ 成立，所以 DPRP-WLS 算法有和定理 3 相同的收敛性结果。

3 数值实验

在本节，对 DPRP-WLS 算法用 Matlab 7.0 编程在 PC (AMD AthlonXP2500+ CPU 1.84GHz) 上进行数值实验。其中 $\delta = 10^{-2}$ ， $\sigma = 0.1$ ， $\alpha_{\min} = 10^{-30}$ ， $\lambda = 0.5$ ， $\xi = \epsilon = 10^{-6}$ ，公式 (8) 中的控制参数 C 分别取 $C = 0$ 和 $C = 1/2$ 进行测试，其中 $C = 0$ 时则对应 PRP 方法。

所测试的函数来自文 [11]，每个函数都选取两个不同的变量维数 n ($50 \leq n \leq 1000$) 来进行测试。限制外循环的最大次数为 10 000 次。每次执行 Wolfe 线搜索前都用下式计算初始的步长 [12]

$$\alpha_0 = \alpha_{k-1} \frac{\nabla f_{k-1}^T \tilde{d}_{k-1}}{\nabla f_{k-1}^T \tilde{d}_{k-1}}, \quad \forall k > 1 \tag{18}$$

当 $k = 1$ 时，选取 $\alpha_0 = \|\nabla f(\tilde{x})\|^{-1}$ 。因为 PRP 方法不能保证算法的下降性，当算法产生一个上升的方向时则用最速下降方向取代当前搜索方向重新开始。具体的数值结果以 NI/NF/NG/NA/Nr 的形式列在表 1 中，其中 NI 表示迭代的次数；NF 表示计算函数值的次数；NG 表示计算函数梯度值的次数；NA 表示应用 ALS 线搜索的次数；Nr 表示搜索方向重新开始的次数；“—”表示方法对这个数值例子失效；CPU time (s) 表示测试所有的这些函数所占用的 CPU 时间 (单位是秒)。

表 1 带不同参数的 DPRP 方法的数值结果 (NI/NF/NG/NA/Nr)

Tab 1 Numerical results of DPRP for different parameters

Problem	Function Name	n	C=0	C=0.5
MGH21	Extended Rosenbrock	500	38/ 116/68/0/2	43/ 148/74/1/0
		1 000	24/ 84/37/0/3	34/ 100/67/0/0
MGH22	Extended Powell Singular	500	369/ 758/571/0/2	56/ 126/90/0/0
		1 000	624/1263/951/0/2	127/ 245/184/0/0
MGH23	Pena101	500	27/ 109/77/2/1	38/128/99/0/0
		1 000	33/ 117/91/0/5	35/ 132/93/0/0
MGH24	Pena102	50	—	1286/ 2595/1963/0/0
		100	116/ 279/188/0/14	134/ 338/220/0/0
MGH25	Variable Dimensioned	100	12/ 56/45/0/0	12/ 56/45/0/0
		200	13/ 62/50/0/0	13/ 62/50/0/0

(接表 1)

Problem	Function Name	n	C=0	C=0.5
MGH26	Trigonometric	100	60/ 157/92/0/5	70/ 146/89/0/0
		200	67/ 139/95/1/3	83/ 208/131/4/0
MGH28	Discrete Boundary Value	500	232/ 383/324/0/11	293/ 397/358/0/0
		1 000	42/ 69/57/0/1	54/ 67/60/0/0
MGH29	Discrete Integral Equation	500	7/ 15/9/0/0	7/ 15/9/0/0
		1 000	7/ 15/9/0/0	8/17/10/0/0
MGH30	Broyden Triagonal	500	35/ 72/46/0/3	32/ 66/40/0/0
		1 000	35/ 68/45/0/6	35/ 70/42/0/0
MGH31	Broyden Banded	500	19/ 76/34/0/0	19/ 56/29/0/0
		1 000	18/ 66/30/0/0	20/ 49/30/0/0
MGH32	Linear Full Rank	500	1/ 3/3/0/0	1/ 3/3/0/0
		1 000	1/ 3/3/0/0	1/ 3/3/0/0
CPU time/s			133.17	94.41

从数值计算的结果来看，除 MGH22和 MGH24外，DPRP方法和 IRP方法的计算效率是相当的。但是，DPRP方法不仅具有良好的收敛性理论，而且因为算法具有充分下降性，数值表现更稳定；而 PRP方法还需要借助重开始策略，才能保证算法被有效地实施。

参考文献:

[1] POLAK E, RIBIERE G. Note sur la convergence de directions conjuguées[J]. Rev Française Informat Recherche Opérationnelle, 1969, 16(3): 35—43.

[2] POLYAK B T. The conjugate gradient method in extreme problems[J]. USSR Comp Math and Math Phys, 1969, 9: 94—112.

[3] ALBAANI M. Descent property and global convergence of the Fletcher—Reeves method with inexact line search[J]. MA J Numer Anal, 1985, 5: 121—124.

[4] GILBERT J C, NOCEDAL J. Global convergence properties of conjugate gradient methods for optimization[J]. SIAM Journal of Optimization, 1992, 2(1): 21—42.

[5] GRIPPO L, LUCIDI S. A globally convergent version of the Polak—Ribiere conjugate gradient method[J]. Mathematics Programming, 1997, 78: 375—391.

[6] GRIPPO L, LUCIDI S. Convergence conditions line

search algorithms and trust region implementations for the Polak—Ribiere conjugate gradient method[J]. Optimization Methods and Software, 2005, 20(1): 71—98.

[7] POWELL M J D. Nonconvex minimization calculations and the conjugate gradient method[C] // Lecture Notes in Mathematics. Berlin: Springer-Verlag, 1984: 1066—122—141.

[8] DAI Y. Analyses of conjugate gradient methods[D]. Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Science (in Chinese), 1997.

[9] POWELL M J D. Convergence properties of algorithms for nonlinear optimization[J]. SIAM Review, 1986, 28: 487—500.

[10] DAI Y. Conjugate gradient methods with Armijo—type line searches[J]. Acta Mathematica Applicata Sinica English Series, 2002, 18(1): 123—130.

[11] MORE J J, GARBOW B S, HILLSTROME K E. Testing unconstrained optimization software[J]. ACM Trans Math Software, 1981, 7: 17—41.

[12] NOCEDAL J, WRIGHT S J. Numerical optimization[M]. Springer Series in Operations Research. New York: Springer, 1999.

(下转第 18 页)

参考文献:

[1] SMITH H L. Monotone Dynamical Systems: an Introduction to the Theory of Competitive and Cooperative Systems [G] // Amer Math Soc Mathematical Surveys and Monographs Vol.41 Providence, R.I. 1995: 48—53.

[2] DEANGELIS D L, POST W M, TRAVIS C C. Positive feedback in natural systems [J]. Biomat 15 Berlin: Springer-Verlag 1986.

[3] HIRSCH M W. Systems of differential equations which are competitive or cooperative. I. Limit sets [J]. SIAM J Math Anal 1982 19(2): 167—179.

[4] 莫宗坚. 代数学 [M]. 北京: 清华大学出版社, 1986.

[5] DEON. Graph theory with applications to engineering and computer science [M]. Englewood Cliffs, N.J.: Prentice-Hall, Inc. 1974.

[6] 北京大学数学系几何与代数教研室代数小组. 高等代数 [M]. 北京: 高等教育出版社, 1988.

A Proof to the Cooperative or Competitive Systems with Respect to Cone

CHEN Ren-lian¹, ZHOU Zuo-ling²

(1. School of Mathematics and Computational Science, Sun Yat-sen University, Guangzhou 510275, China;
2. Lingnan College, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: A proof is given using graph theory by induction. Whether a system is cooperative or competitive in a domain D with respect to cone K_n is equivalent to for every closed loop in G , the number of edges with “—” sign is even or odd.

Key words: cone K_n ; cooperative system; competitive system

(上接第 14 页)

Modified PRP Methods with Sufficient Descent Property and Their Convergence Properties

YU Gao-hang¹, GUAN Li-tai²

(Department of Scientific Computation and Computer Applications,
Sun Yat-sen University, Guangzhou 510275, China)

Abstract: The conjugate gradient (CG) method is very suitable for solving large-scale nonlinear optimization due to the simplicity of their iteration and their very low memory requirements. Generally, the PRP method is regarded as one of the most effective CG methods. However, the PRP method even with the exact line search is not globally convergent on nonconvex functions. Based on a modified PRP formula, a class of conjugate gradient methods were proposed for unconstrained optimization. No line search is required, these methods possess the sufficient descent property. Under suitable conditions, globally convergent results were given in this paper. Preliminary numerical results show that these methods are effective.

Key words: unconstrained optimization; large-scale optimization; conjugate gradient method; global convergence