

新线搜索下修正 PRP 共轭梯度法的全局收敛性及其数值结果 *

Global Convergence and Numerical Results of a Modified PRP Conjugate Gradient Method with a New Line Search

王松华¹,吴加其²

WANG Songhua¹,WU Jiaqi²

(1.百色学院数学与统计学院,广西百色 533000;2 广西大学数学科学学院,广西南宁 530004)

(1. College of Mathematics and Statistics Science, Baise University, Baise, Guangxi, 533000, China;2. College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, 530004, China)

摘要:针对大规模非线性无约束问题,采用文献[9]提出的新型线搜索和文献[10]修正 PRP 公式设计一个新的算法。在适当的条件下,证明新算法具有全局收敛性。初步的数值试验结果表明,新算法是有效的,适合求解大规模非线性无约束优化问题。

关键词:非线性无约束优化 共轭梯度法 线搜索 全局收敛性

中图分类号:O224 **文献标识码:**A **文章编号:**1005-9164(2018)06-0728-06

Abstract: Aiming at the large-scale nonlinear unconstrained problem, a new algorithm is designed by using the new line search proposed in the literature [9] and a modified PRP formula proposed in the literature [10]. Under appropriate conditions, the new algorithm is proved to have global convergence. Preliminary numerical test results show that the new algorithm is effective and suitable to solve some large-scale nonlinear unconstrained optimizations.

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

0 引言

考虑非线性无约束最优化问题:

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

其中目标函数 $f(x):R^n \rightarrow R$ 是连续可微非线性函数。共轭梯度法在计算过程中只需运用一阶导数信息,存储信息少,其收敛速度介于最速下降法和牛顿法之间,具有二次终止性,在求解非线性无约束优化问题时有良好的数值效果,而且编程简便,是最优化问题求解中最常用的方法之一,其迭代公式为

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

x_{k+1} 和 x_k 分别是第 $k + 1$ 次和第 k 次迭代点, α_k 是步

收稿日期:2018-08-29

作者简介:王松华(1970—),男,讲师,主要从事最优化理论与方法的研究,E-mail:523429892@qq.com。

* 广西自然科学基金项目(ZD2014132)和广西教育厅科研项目(YB2014389, YB2014381)资助。

长因子, d_k 是搜索方向, 其迭代公式定义为

$$d_{k+1} = \begin{cases} -g_{k+1} & \text{if } k=0 \\ -g_{k+1} + \beta_k d_k & \text{if } k \geq 1 \end{cases}, \quad (2)$$

其中参数 $\beta_k \in R$, g_{k+1} 是函数 $f(x)$ 在点 x_{k+1} 上的梯度函数。共轭梯度法主要考虑步长因子 α_k 和搜索方向 d_k 两个因素。众所周知, 步长因子 α_k 计算有精确线搜索和非精确线搜索两种方法, 因为精确线搜索在计算过程中计算量太大, 因此在实际计算中主要采用非精确线搜索。经典非精确线搜索方法有以下几种: 弱 Wolfe-Powell(WWP)线搜索^[1], 强 Wolfe-Powell(SWP)线搜索^[2], 推广 Wolfe-Powell 线搜索^[3], Armijo 线搜索^[4] 和 Grippo-Lucidi 线搜索^[5] 等。研究表明, 非精确线搜索的建立, 需要考虑到两个问题^[6]: 满足非精确线搜索准则的 α_k 是否存在, 采用非精确线搜索准则的共轭梯度法是否全局收敛。在实际运算中, 线搜索需要一定量的运算时间, 而且线搜索的选择会影响到算法的计算效果。近来, 针对于 PRP 共轭梯度法的线搜索研究取得了一些成果, 较为典型的有韦增欣的 3 种新型线搜索^[7-9], 也已经得到广泛使用。

2017 年, Yuan 等^[10] 在其论著中提出一种新型线搜索(简称 MWWP 型线搜索), 在较强的条件下该线搜索对经典 PRP 公式具有全局收敛的性质, 数值实验结果表现很好, 更具竞争性。MWWP 型线搜索形式如下:

$$f(x_k + \alpha_k d_k) \leq f_k + \delta \alpha_k g_k^T d_k + \alpha_k \min[-\delta_1 g_k^T d_k, \delta \frac{\alpha_k}{2} \|d_k\|^2] \quad (3)$$

和

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k + \min[-\delta_1 g_k^T d_k, \delta \alpha_k \|d_k\|^2], \quad (4)$$

其中, $\delta \in (0, 1/2)$, $\delta_1 \in (0, \delta)$ 和 $\sigma \in (\delta, 1)$ 。

同样的, 针对公式(2) 参数 β_k 选取不同的公式即为不同的方法, 经典的有以下 6 种: FR、PRP、HS、LS、DY 和 CD 等方法^[11]。这些个方法各具特点, PRP、HS 和 LS 方法数值好但在一般条件下不具备全局收敛性, FR、DY 和 CD 方法理论上收敛性表现良好但是数值结果不是很理想, 总体而言在数值结果上 PRP 方法表现最为优越。针对 PRP 共轭梯度法收敛性质差的缺陷, 许多研究者通过修正 PRP 公式, 在适当的线搜索上设计新的算法, 取得丰富的成果。其中, 黎勇^[12] 给出一种修正 PRP 共轭梯度法, 其参数 β_k^* 形式如下:

$$\beta_k^* = \frac{g_k^T (\frac{\|g_{k-1}\|}{\|g_k\|} g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}, \quad (5)$$

式(5) 参数 β_k^* 具有非负性等性质, 通过式(5)建立的共轭梯度法均具有较好的全局收敛性^[12-14], 在数值试验上针对选取的测试函数计算出的结果普遍良好且稳定^[12-13], 适合求解实际大规模非线性无约束最优化问题。

但是, 式(5)在 MWWP 型线搜索下是否也具有全局收敛性质, 数值结果表现如何等情况尚未提出。一个好的共轭梯度法通常与选定的线搜索有很大的关系, 基于此, 本文利用 MWWP 型线搜索和公式(5)设计一个新的算法, 分析其全局收敛性, 并分别采用其他经典的几类线搜索对公式(5)进行大规模问题的数值实验, 分析新算法的有效性, 最终通过对 MWWP 型线搜索和公式(5)的进一步研究, 给大规模非线性无约束优化问题提供更好的解决方法。

1 新线搜索下修正 PRP 共轭梯度法

为方便表述, 本文将新算法称为算法 1, 其步骤如下。

步骤 1: 选择一个初始点 $x_1 \in R^n$, $\epsilon \in (0, 1)$, $\delta \in (0, 1/2)$, $\delta_1 \in (0, \delta)$, $\sigma \in (\delta, 1)$ 。令 $d_1 = -g_1 = -\nabla f(x_1)$, $k=1$;

步骤 2: 如果 $\|g_k\| \leq \epsilon$, 则停止;

步骤 3: 利用线搜索(4) 和(5) 计算步长 α_k ;

步骤 4: 令 $x_{k+1} = x_k + \alpha_k d_k$;

步骤 5: 如果 $\|g_{k+1}\| \leq \epsilon$, 则停止;

步骤 6: 利用公式(3) 计算参数 β_k^* , 公式(2) 计算搜索方向 $d_{k+1} = -g_{k+1} + \beta_k^* d_k$;

步骤 7: 令 $k := k + 1$, 然后继续步骤 2。

一般的, 在非线性共轭梯度法收敛性的证明中需要用到以下假设。

假设 1:

(A) 水平集 $L_0 = \{x \mid f(x) \leq f(x_0)\}$ 有界, 其中 $f(x_0)$ 是初始点。

(B) 令 $f(x)$ 是连续可微, 且梯度函数 $g(x)$ 在水平集 $L_0 > 0$ 内 Lipschitz 连续, 即

$$\|g(x) - g(y)\| \leq L \|x - y\|, x, y \in R^n. \quad (6)$$

因为 $\{f(x_k)\}$ 是单调递减序列, 因此由算法产生的序列 $\{x_k\}$ 包含在水平集 L_0 中, 且存在一个常数 $X_M > 0$, 使得

$$\|x\| \leq X_M, x \in L_0. \quad (7)$$

2 全局收敛性分析

Yuan 等^[10] 指出, MWWP 线搜索有两种情况:

$$\textcircled{1} \min(-\delta_1 g(x_k)^T d_k, \delta \alpha_k \|d_k\|^2) =$$

$$\delta\alpha_k \|d_k\|^2;$$

$$\textcircled{2} \min(-\delta_1 g(x_k)^T d_k, \delta\alpha_k \|d_k\|^2) =$$

$$-\delta_1 g(x_k)^T d_k$$

本文主要讨论算法 1 在 MWWP 线搜索情况①条件下的全局收敛性。

为证明算法 1 的全局收敛性,下面给出假设 2。

假设 2:下面的关系式

$$d_k^T g_k < 0 \text{ 和 } g_{k+1} d_k \leq -\sigma_1 g_k^T d_k \quad (8)$$

对任意的 k 都成立,其中 $\sigma_1 \leq \sigma$ 。

以下我们将讨论新算法的全局收敛性。首先给出引理 1 和引理 2,其具体的证明详见 Yuan 等^[10]的 Lemma 4.1 和 Lemma 4.2,本文不再赘述。

引理 1 考虑算法 1,且假设 1 和假设 2 成立,如果序列 $\{x_k, \alpha_k, d_k, g_k\}$ 由算法 1 生成,则下面两个式子必有一个成立:

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty \quad (9)$$

或

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (10)$$

引理 2 假设条件满足引理 1,如果

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} < \infty, \quad (11)$$

那么下式

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0 \quad (12)$$

成立。

定理 1 假设条件满足引理 1,且 MWWP 线搜索满足情况①,即

$$\min(-\delta_1 g(x_k)^T d_k, \delta\alpha_k \|d_k\|^2) = \delta\alpha_k \|d_k\|^2,$$

那么有

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0.$$

证明:采用反证法来证明。假设定理不成立,那么存在一个常数 $\epsilon_0 > 0$,使得

$$\|g_k\| \geq \epsilon_0, \forall k \geq 1.$$

由假设 1,则存在一个常数 $G_M > 0$,使得 $\|g(x)\| \leq G_M, x \in R^n$,且 $d_{k+1} = -g_{k+1} + \beta_k d_k$ 。于是有

$$\begin{aligned} \|d_{k+1}\| &\leq \| -g_{k+1} + \beta_k^* d_k \| \leq \|g_{k+1}\| + \|\beta_k^*\| \|d_k\| \leq \|g_{k+1}\| + \\ &\frac{\|g_{k+1}^T\| \|g_k\|}{\|g_k\|^2} \|g_{k+1} - g_k\| \|d_k\| = \|g_{k+1}\| + \\ &\frac{\|g_{k+1}^T\| \|g_k\|}{\|g_{k+1}\|} \|g_{k+1} - g_{k+1} + g_{k+1} - g_k\| \cdot \end{aligned}$$

$$\|d_k\| \leq \|g_{k+1}\| +$$

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

$$\|d_k\| \leq \|g_{k+1}\| + \frac{\|g_{k+1}^T\| \|g_{k+1} - g_k\| \|d_k\|}{\|g_k\|^2} \leq$$

$$\|g_{k+1}\| + \frac{2G_M L \|s_k\| \|d_k\|}{\|g_k\|^2} \leq \|g_{k+1}\| + \frac{2G_M L \|s_k\|}{\epsilon_0^2} \|d_k\|,$$

又由式(3),得到

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta\alpha_k g_k^T d_k -$$

$$\alpha_k \min[-\delta_1 g_k^T d_k, \delta \frac{\alpha_k}{2} \|d_k\|^2] \geq -(\delta - \delta_1) g_k^T d_k.$$

对上述不等式从 $k=0$ 到 ∞ 求和,结合假设 1(A),得到

$$\sum_{k=0}^{\infty} (-\delta_1 \alpha_k g_k^T d_k) < \infty. \quad (13)$$

由 MWWP 线搜索的情况①: $\min(-\delta_1 g(x_k)^T d_k, \delta\alpha_k \|d_k\|^2) = \delta\alpha_k \|d_k\|^2$ 和式(13),有

$$\sum_{k=0}^{\infty} (\delta \|s_k\|^2) = \sum_{k=0}^{\infty} (\delta\alpha_k^2 \|d_k\|^2) \leq$$

$$\sum_{k=0}^{\infty} (-\delta_1 \alpha_k g_k^T d_k) < \infty.$$

则

$$\|s_k\| \rightarrow \infty, k \rightarrow \infty.$$

从而,存在一个整数 $k_0 \geq 0$ 和一个常数 $s \in (0, 1)$,对所有的 $k \geq k_0$,使得

$$\frac{2G_M L \|s_k\|}{\epsilon_0^2} \leq s.$$

因此,对所有的 $k \geq k_0$,我们有

$$\|d_{k+1}\| \leq G_M + s \|d_k\| \leq G_M(1 + s + s^2 + \dots + s^{k-k_0-1}) + s^{k-k_0} \|d_{k_0}\| \leq \frac{G_M}{1-s} + \|d_{k_0}\|.$$

令 $D_M = \max\{\|d_1\|, \|d_2\|, \dots, \|d_{k_0}\|, \frac{G_M}{1-s} + \|d_0\|\}$,我们得到

$$\|d_k\| \leq G_M, \forall k > 0.$$

这与引理 2 矛盾。证明完毕。

3 数值实验和新算法有效性分析

为考查本文提出的新搜索下修正 PRP 共轭梯度法方法(算法 1)的数值表现,下面我们将其与算法 2、算法 3 和算法 4 进行比较。相关的参数选取分别如下。

算法 1:公式(3)与 MWWP 线搜索^[10],其中 $\delta = 0.49, \delta_1 = 0.24, \sigma = 0.67$;

算法 2:公式(3)与弱 Wolfe-Powell 线搜索^[1],其

中 $\delta = 0.49, \sigma = 0.67$;

算法 3: 公式(3)与广义 Wolfe-Powell 线搜索^[3], 其中 $\delta = 0.49, \sigma_1 = 0.67, \sigma_2 = 11.12$;

算法 4: 公式(3)与强 Wolfe-Powell 线搜索^[2], 其中 $\delta = 0.49, \sigma = 0.67$.

此次数值实验, 选用 Bongartz 等^[15]和 Andrei^[16]提出的 CUTEr 数据库中 69 个测试问题, 分别对 4

表 1 测试问题

No.	问题 Problem	No.	问题 Problem
1	Extended Freudenstein and Roth Function	36	BDQRTIC Function (CUTE)
2	Extended Trigonometric Function	37	TRIDIA Function (CUTE)
3	Extended Rosenbrock Function	38	ARWHEAD Function (CUTE)
4	Extended White and Holst Function	39	ARWHEAD Function (CUTE)
5	Extended Beale Function	40	NONDQUAR Function (CUTE)
6	Extended Penalty Function	41	DQDRTIC Function (CUTE)
7	Perturbed Quadratic Function	42	EG2 Function (CUTE)
8	Raydan 1 Function	43	DIXMAANA Function (CUTE)
9	Raydan 2 Function	44	DIXMAANB Function (CUTE)
10	Diagonal 1 Function	45	DIXMAANC Function (CUTE)
11	Diagonal 2 Function	46	DIXMAANE Function (CUTE)
12	Diagonal 3 Function	47	Broyden Tridiagonal Function
13	Hager Function	48	Almost Perturbed Quadratic Function
14	Generalized Tridiagonal 1 Function	49	Tridiagonal Perturbed Quadratic Function
15	Extended Tridiagonal 1 Function	50	EDENSCH Function (CUTE)
16	Extended Three Exponential Terms Function	51	VARDIM Function (CUTE)
17	Generalized Tridiagonal 2 Function	52	STAIRCASE S1 Function
18	Diagonal 4 Function	53	LIARWHD Function (CUTE)
19	Diagonal 5 Function	54	DIAGONAL 6 Function
20	Extended Himmelblau Function	55	DIXON3DQ Function (CUTE)
21	Generalized PSC1 Function	56	DIXMAANF Function (CUTE)
22	Extended PSC1 Function	57	DIXMAANG Function (CUTE)
23	Extended Powell Function	58	DIXMAANH Function (CUTE)
24	Extended Block Diagonal BD1 Function	59	DIXMAANI Function (CUTE)
25	Extended Maratos Function	60	DIXMAANJ Function (CUTE)
26	Extended Cliff Function	61	DIXMAANK Function (CUTE)
27	Quadratic Diagonal Perturbed Function	62	DIXMAANL Function (CUTE)
28	Extended Wood Function	63	DIXMAAND Function (CUTE)
29	Extended Hiebert Function	64	ENGVAL1 Function (CUTE)
30	Quadratic Function QF1 Function	65	FLETCHCR Function (CUTE)
31	Extended Quadratic Penalty QP1 Function	66	COSINE Function (CUTE)
32	Extended Quadratic Penalty QP2 Function	67	Extended DENSCHNB Function (CUTE)
33	A Quadratic Function QF2 Function	68	DENSCHNF Function (CUTE)
34	Extended EP1 Function	69	SINQUAD Function (CUTE)
35	Extended Tridiagonal-2 Function		

我们采用 Dolan 和 Moré^[17]提出的比较分析方法。根据本次实验的数值结果, 针对 4 种算法的迭代次数(NI)、目标函数值计算次数(NF)、运行时间(CPUtime)和目标函数梯度计算次数(NG)等 4 个指标, 分别作出算法 1、算法 2、算法 3 和算法 4 的性能

种算法进行模拟试验。电脑系统为 Intel(R) Xeon (R)CPU, E5507 @ 2.27 GHz, 4.00 GB 内存, Windows 7 操作系统。采用 MATLAB2014b 编程和运行, 维数取值为 [4500, 9000, 15000, 45000], 终止条件: $\|g_k\| \leq 10^{-5}$ 或迭代次数超过 800。测试问题见表 1。

比较图, 详见图 1~4。从性能比较图的曲线变化走向可以看出, 在相同的运行环境下, 对选取的 69 个测试问题, 算法 1 在迭代次数、目标函数值的计算次数、运行时间和目标函数梯度的计算次数上, 都比算法 2、算法 3 和算法 4 的计算效率更好, 稳定性更优。

4 结语

本文针对大规模非线性无约束优化问题,在经典非精确线搜索的基础上,利用 Yuan 等^[10]提出的新型线搜索技术,结合黎勇^[12]的修正 PRP 公式,设计出一种新的 PRP 共轭梯度法。在适当的条件下,证明新算法的全局收敛性。对选取的 69 个测试函数进行初步的数值试验,维数取值范围在 [4500, 9000, 15000, 45000],并与 3 种经典非精确线搜索技术下的算法进行比较分析。结果表明,新算法能够有效求解 45000 维的大规模问题,总体表现更好,新算法是有效的。

在未来的工作中,我们将对新线索的理论进行思考和讨论,如新线索在其他共轭梯度法上的应用,新算法中相关参数选取对算法运行效果的影响;同时,新算法在解决实际大规模非线性无约束优化问题的效果有待进一步验证。

参考文献:

- [1] MCCORMICK G P. A modification of Armijo's step-size rule for negative curvature[J]. *Mathematical Programming*, 1977, 13(1): 111-115.
- [2] YUAN Y X, SUN W Y. *Theory and methods of optimization*[M]. Beijing: Science Press of China, 1999.
- [3] 李董辉,童小娇,万中. *数值最优化*[M]. 北京: 科学出版社, 2005.
LI D H, TONG X J, WAN Z. *Numerical optimization* [M]. Beijing: Science Press, 2005.
- [4] ARMIJO L. Minimization of functions having lipschitz conditions partial derivatives [J]. *Pacific Journal of Mathematics*, 1966, 16(1): 1-3.
- [5] GRIPPO L, LAMPARIELLO F, LUCIDI S. A nonmonotone line search technique for Newton's methods[J]. *SIAM J Numer Anal*, 1986, 23(4): 707-716.
- [6] 高立. *数值最优化方法*[M]. 北京: 北京大学出版社, 2014.
GAO L. *Numerical optimization method*[M]. Beijing: Peking University Press, 2014.
- [7] WEI Z X, LI G Y, QI L Q. Global convergence of the Polak-Ribière-Polyak conjugate gradient method with an Armijo-type inexact line search for nonconvex unconstrained optimization problems [J]. *Mathematics of Computation*, 2008, 77: 2173-2193.
- [8] WEI Z X, QI L Q, CHEN X. An SQP-type method and its application in stochastic programs[J]. *Journal of Optimization Theory and Applications*, 2003, 116(1): 205-228.
- [9] WEI Z X, QI L Q, ITO S. New step-size rules for opti-

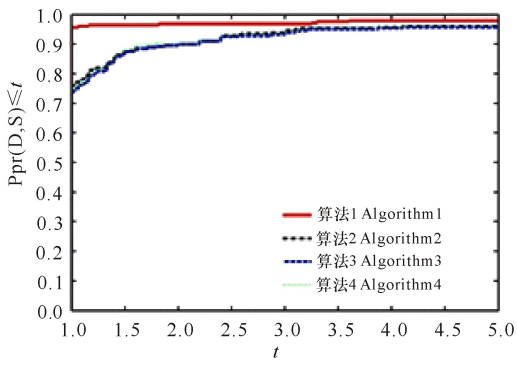


图 1 迭代次数比较图

Fig. 1 Comparison graph of iteration number

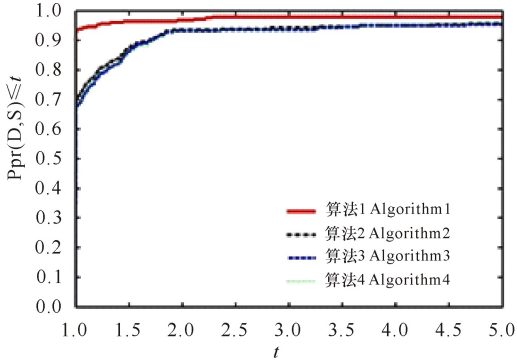


图 2 目标函数值计算次数比较图

Fig. 2 Comparison graph of calculation number of objective function value

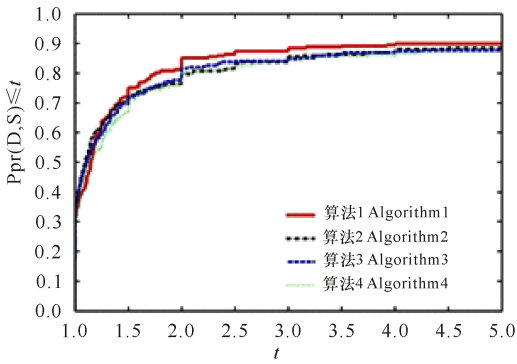


图 3 运行时间比较图

Fig. 3 Comparison graph of running time

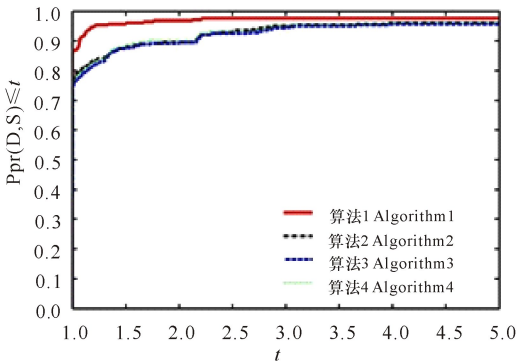


图 4 目标函数梯度值计算次数比较图

Fig. 4 Comparison graph of calculation number of objective function gradient value

- mization problems[J]. International Conference on Recent Advances in Computational Mathematics, Japan SIAM, 2001.
- [10] YUAN G L, WEI Z X, LU X W. Global convergence of BFGS and PRP methods under a modified weak Wolfe-Powell line search[J]. Applied Mathematical Modelling, 2017, 47: 811-825.
- [11] 戴或虹, 袁亚湘. 非线性共轭梯度法[M]. 上海: 上海科学技术出版社, 2000.
DAI Y H, YUAN Y X. Nonlinear conjugate gradient method[M]. Shanghai: Shanghai Scientific & Technical Publishers, 2000.
- [12] 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.
LI Y. Global convergence and numerical experiments of some new modification of PRP conjugate gradient method[J]. Journal of Southwest University: Natural Science, 2011, 33(11): 23-28.
- [13] 黎勇. 一种修改的 PRP 共轭梯度法[J]. 广西科学, 2013, 20(1): 5-8.
LI Y. A modification PRP conjugate gradient method [J]. Guangxi Sciences, 2013, 20(1): 5-8.
- [14] 黎勇. 无约束优化中的一种修正的 PRP 共轭梯度法的全局收敛性分析[J]. 百色学院学报, 2010, 23(3): 55-59.
LI Y. Global convergence analysis of a modification PRP conjugate gradient method for unconstrained optimization[J]. Journal of Baise University, 2010, 23(3): 55-59.
- [15] BONGARTZ I, CONN A R, GOUNLD N. CUTE: Constrained and unconstrained testing environment [J]. Acm Transactions on Mathematical Software, 1995, 21(1): 123-160.
- [16] ANDREI N. An unconstrained optimization test functions collection [J]. Environmental Science and Technology, 2008, 10(1): 6552-6558.
- [17] DOLAN E D, MORÉ J J. Benchmarking optimization software with performance profiles [J]. Mathematical Programming, 2001, 91(2): 201-213.

(责任编辑: 米慧芝)