

一种求解无约束优化问题的新混合共轭梯度法^①

唐 天 国

南充职业技术学院 电子信息工程系, 四川 南充 637000

摘要: 在现有共轭梯度方法的基础上, 提出一种新混合共轭梯度法来求解无约束最优化问题. 该方法采用近似方法去逼近 Hessian 矩阵, 克服了传统牛顿法求解 Hessian 矩阵中存在的计算量大等问题, 并在强 wolfe 线搜索技术下给出该共轭梯度算法的全局收敛性证明. 实验结果表明, 与 PRP(Polak-Ribiere-Polyak)方法和 HYBRID(混合)方法相比较, 该文提出的新混合共轭梯度算法的迭代时间少于前两者方法, 说明该文方法可行、有效.

关 键 词: 共轭梯度法; 无约束优化; 强 Wolfe 线搜索; 全局收敛性

中图分类号: O221

文献标志码: A

文章编号: 1000-5471(2019)09-0034-06

最优化理论作为一门应用性强的学科, 是运筹学的重要组成部分^[1], 而求解最优化方法的常用方法之一是共轭梯度法^[2], 其在算法收敛速度、存储需求等方面的综合性能要优于最速下降法和牛顿法, 因此共轭梯度法用来求解无约束优化问题^[3], 特别是大规模的优化问题, 在化学、经济和工程设计等领域具有较多应用, 成为应用领域人员的热点研究对象.

对于连续可微函数 $f(x)$, 其无约束优化问题表示为 $\min_{x \in R} f(x)$, 对于求解上述优化的函数共轭梯度法可用如下描述来表示: k 为迭代次数, 确定当前迭代点 x_k 的搜索方向 d_k , 该方向即为目标函数 $f(x_k)$ 的下降方向, 由搜索步长 α_k 得到下一个迭代点 $x_{k+1} = x_k + \alpha_k d_k$. $g_k = \nabla f(x_k)$ 为梯度, 则搜索方向与梯度的关系为

$$d_k = \begin{cases} -g_k & k = 0 \\ -g_k + \beta_k d_{k-1} & k \geq 1 \end{cases} \quad (1)$$

当式(1)中搜索方向确定, 选择一个合适的搜索步长尤为关键, 常见的线搜索方法有: 标准 Armijo 法、Goldstein 步长规则、Wolf 步长规则和强 Wolf 步长规则等.

1. 标准 Armijo 法, 设 $\beta > 0$, $\gamma \in (0, 1)$, 取步长 $\alpha_k = \beta \gamma^{m_k}$, 其中 m_k 是满足下式的最小非负整数

$$f(x_k + \beta \gamma^{m_k} d_k) \leq f(x_k) + \sigma \beta \gamma^{m_k} g_k^T d_k \quad (2)$$

式(2)中 $\sigma \in (0, 1)$.

2. Goldstein 步长规则^[4]

$$\begin{aligned} f(x_k + \alpha d_k) &\leq f(x_k) + \alpha g_k^T d_k \\ f(x_k + \alpha d_k) &> f(x_k) + (1 - \sigma) \alpha g_k^T d_k \end{aligned} \quad (3)$$

式(3)中 $\sigma \in (0, \frac{1}{2})$, 当 $\alpha > 0$ 足够小的时候, 第 2 个式子一定不成立.

3. Wolf 步长规则

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \sigma_1 \alpha_k g_k^T d_k \\ \nabla f(x_k + \alpha_k d_k)^T d_k &\geq \sigma_2 g_k^T d_k \end{aligned} \quad (4)$$

① 收稿日期: 2018-08-14

基金项目: 四川省高等职业教育研究中心 2016 年科研课题 (GZY16B09).

作者简介: 唐天国(1965-), 男, 副教授, 主要从事数学与计算机科学研究.

式(4)中 $0 < \sigma_1 < \sigma_2 < 1, \alpha_k > 0$.

4. 强 Wolf 步长规则

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \sigma_1 \alpha_k g_k^T d_k \\ |\nabla f(x_k + \alpha_k d_k)^T d_k| &\geq \sigma_2 |g_k^T d_k| \end{aligned} \quad (5)$$

式(5)中 $0 < \sigma_1 < \sigma_2 < \frac{1}{2}, \alpha_k > 0$.

共轭梯度算法中 β_k 为一个关键参数, 常见的 β_k 选择方式主要有 HS(Hestenes-Stiefel), PRP, DY 和 CD(Conjugate-Descent) 等方法, 不同的共轭梯度法具有不同的 β_k 表达式, PRP 方法和 DY(Dai-Yuan) 方法^[5-6] 的表达式分别为

$$\beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}; \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})} \quad (6)$$

PRP方法在迭代速度方面具有优势, 劣势就是全局收敛性差; DY方法则相反, 其最大的优点即全局收敛性好, 缺点为对搜索步长的要求较高. 周雪琴^[7] 给出了具有充分下降性的新共轭梯度法及全局收敛性, 给出了一种新的 β_k 表达式, 并证明了在强 Wolfe 搜索条件下的全局收敛性. 何晓旭等^[8] 将 FR(Fletcher-Reeves) 方法与 PRP 方法混合, 新方法具有算法效率高, 所需存储量小, 不需要外来参数等优点, 其公式为

$$\beta_k^{GN} = \max\{\min[-c\beta_k^{PRP}, \beta_k^{FR}], \min[\beta_k^{FR}, \beta_k^{PRP}]\} \quad (7)$$

车彩丽^[9] 中给出了混合 HS-DY 共轭梯度算法, 该算法的优点是具有更好的收敛性, 且数值表现更优越、误差更小. 刘玲等^[10] 提出了一种带参数的混合共轭梯度算法, 该算法的优点是不依赖线搜索的前提下, 使得搜索方向满足充分下降条件, 其缺点是在搜索过程中需要引入重新开始的策略, 降低计算效率. 常见的谱共轭梯度法公式为

$$\beta_k = \frac{(\theta_k y_{k-1} - s_{k-1})^T g_k}{d_{k-1}^T y_{k-1}} \quad (8)$$

本文对以上方法进行综合分析, 并在此基础上提出了一种新混合共轭梯度法公式, 该方法采用强 Wolf 步长搜索法则, 解决了计算量大的问题, 提高了计算效率, 且具有全局收敛性和较好的数值特性.

1 本文新混合共轭梯度算法

结合 PRP 方法、DY 方法、BFGS 方法的优点, 本文提出一种改进的混合共轭梯度法

$$d_0 = -g_0, d_{k+1} = -Q_{k+1} g_{k+1}, k = 0, 1, 2, \dots \quad (9)$$

式(9)中 Q_{k+1} 是一个 $N \times N$ 的矩阵, 表达式为

$$\begin{aligned} Q_{k+1} &= \theta_{k+1} I - \theta_{k+1} \frac{z_k s_k^T + s_k z_k^T}{s_k^T z_k} + \\ &\quad \left(1 + \theta_{k+1} \frac{z_k^T z_k}{s_k^T z_k}\right) \frac{s_k s_k^T}{s_k^T z_k} \end{aligned} \quad (10)$$

式(10)中 $s_k = d_k - d_{k-1}, z_k = g_k - g_{k-1}$. θ_{k+1} 是典型方法中常用的搜索方向, 由 BFGS 算法及搜索思想, 可以得到式(10)中本文提出的新混合共轭梯度法公式. 对于 Hessen 矩阵, 该算法用近似方法直接取代直接求解方法, 在强 Wolfe 线搜索条件下, 可以得到本文算法具有如下性质

$$s_k^T y_k = s_k^T g_{k+1} - s_k^T g_k \geq -(1 - \sigma) s_k^T g_k > 0 \quad (11)$$

图1是改进的混合共轭梯度法算法流程图, 其中 k 为迭代步数, 从式(9)、式(10)中可以得到 β_k , 式(5)中是强 Wolf 步长规则.

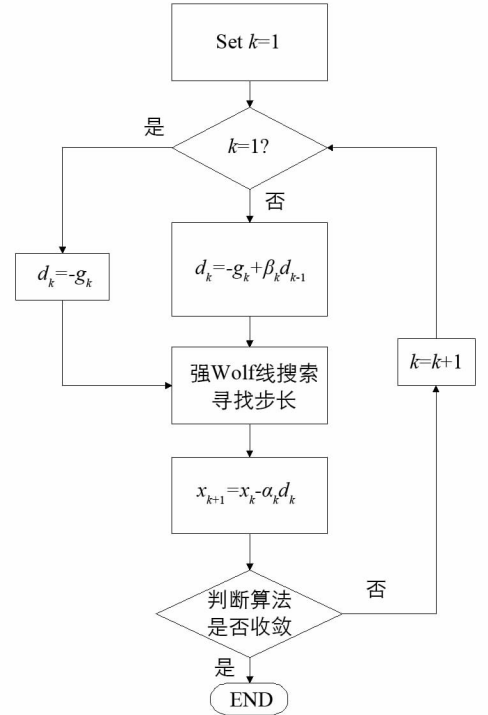


图1 本文新混合共轭梯度算法流程图

对于本文算法, 如果满足 $|f(x_{k+1}) - f(x_k)| \leq \epsilon$, 则算法终止迭代, 其中, ϵ 是较小的正数, ϵ 的选取一般由经验人为给定.

2 全局收敛性证明

对于全局收敛性, 如果满足如下条件, 则说明算法具有全局收敛性. 满足条件: 对于任意初始点, 算法开始并产生的迭代能够收敛到最优. 本文采用反证法来进行收敛性证明, 具体描述如式(12).

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

如果满足上式, 则认为算法为弱收敛, 即满足条件: 只有目标函数导数下界的极限需要接近零. 在这里引入 2 个引理:

引理 1 对该混合共轭梯度算法, 步长选择法采用强 Wolfe 规则, 则有如下表达式成立

$$g_k^T d_k < 0 \quad (13)$$

证 由于 $\varphi_k(\alpha) = f(x_k + \alpha d_k)$ 在 $\alpha > 0$ 时存在下界, 故射线 $\varphi(\alpha) = f(x_k) + \sigma_1 \alpha g_k^T d_k$ 与 $\varphi_k(\alpha)$ 在正半轴上有交点, 令该最小交点为 α' , 则有以下表达式成立

$$f(x_k + \alpha' d_k) = f(x_k) + \sigma_1 \alpha' g_k^T d_k \quad (14)$$

利用中值定理可得, 存在 $\alpha'' \in (0, \alpha')$ 使得式(15)成立

$$f(x_k + \alpha' d_k) - f(x_k) = \alpha' \nabla f(x_k + \alpha'' d_k)^T d_k \quad (15)$$

根据强 Wolfe 步长搜索条件, 由式(14)和式(15)可得

$$\nabla f(x_k + \alpha'' d_k)^T d_k = \sigma_1 g_k^T d_k > \sigma_2 g_k^T d_k \quad (16)$$

由此可得 $g_k^T d_k < 0$, 引理 1 得证.

引理 2 对于任意的 k , 有以下不等式成立

$$-\sum_{i=0}^k \sigma_i^2 \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -2 + \sum_{i=0}^k \sigma_i^2 \quad (17)$$

由数据归纳法可知, 当 $k=0$ 时 $d_0 = -g_0$, 上式显然成立; 先需要验证在 $k > 0$ 时该不等式也成立, 则通过 FR 计算公式有

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = -1 + \frac{g_k^T d_k}{\|g_k\|^2} \quad (18)$$

由引理 1 可得

$$-1 + \sigma_2 \frac{g_k^T d_k}{\|g_k\|^2} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 - \sigma_2 \frac{g_k^T d_k}{\|g_k\|^2} \quad (19)$$

利用归纳假设对上式中不等式右侧的项进行处理可得

$$-\sum_{j=0}^{k+1} \sigma_j^2 = -1 - \sigma_2 \sum_{j=0}^k \sigma_j^2 \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 + \sigma_2 \sum_{j=0}^k \sigma_j^2 = -2 + \sum_{j=0}^{k+1} \sigma_j^2 \quad (20)$$

由此, 引理 2 得证. 从引理 2 我们可得如下不等式

$$-\frac{1}{1 - \sigma_2} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq \frac{2\sigma_2 - 1}{1 - \sigma_2} \quad (21)$$

在全局收敛性的证明中还需要满足 2 个条件.

1. 目标函数有界

$$L = \{x \in R^n \mid f(x) \leq f(x_0)\} \quad (22)$$

2. 函数 $f(x)$ 连续可微, 其导函数 $g(x)$ 是 Lipschitz 连续的

$$\|g(x_1) - g(x_2)\| \leq L \|x_1 - x_2\| \quad (23)$$

首先由式(5)中强 Wolf 线性搜索步长规则可以得出

$$|g_k^T d_{k-1}| \leq -\sigma_2 g_{k-1}^T d_{k-1} \leq \frac{\sigma_2}{1 - \sigma_2} \|g_{k-1}\|^2 \quad (24)$$

结合引理 1 与 β_{k-1} 的公式, 可得 d_k 的递推表达式

$$\begin{aligned}\|d_k\|^2 &= \|g_k\|^2 - 2\beta_{k-1}g_k^T d_{k-1} + \beta_{k-1}^2 \|d_{k-1}\|^2 \leqslant \\ &\|g_k\|^2 - \frac{2\sigma_2}{1-\sigma_2} \|g_k\|^2 + \beta_{k-1}^2 \|d_{k-1}\|^2 = \\ &\frac{1+\sigma_2}{1-\sigma_2} \|g_k\|^2 + \beta_{k-1}^2 \|d_{k-1}\|^2\end{aligned}\quad (25)$$

将式(25) 中的 $\|d_{k-1}\|^2$ 进行代换, 由此可得反复的迭代不等式

$$\|d_k\|^2 \leqslant \frac{1+\sigma_2}{1-\sigma_2} \|g_k\|^4 \sum_{i=1}^k \frac{1}{\|g_i\|^2} + \frac{\|g_k\|^2}{\|g_0\|^2} \quad (26)$$

根据式(12), 当满足式(12) 中条件则认为该算法收敛, 使用反证法来证明本文算法具有收敛性, 首先假设不收敛, 则存在一常数 $\epsilon > 0$, 使得式(27) 成立

$$\|g_k\| \geqslant \epsilon \quad (27)$$

根据假设 1 可知 g_k 在水平集上有界, 则对于 r, ϵ 满足式(28) 中表达式

$$\frac{1}{r^2} \leqslant \frac{1}{\|g_k\|} \leqslant \frac{1}{\epsilon^2} \quad (28)$$

将式(28) 对式(26) 进行不等式放缩, 得到式(29) 成立

$$\|d_k\|^2 \leqslant \frac{1+\sigma_2}{1-\sigma_2} \|r\|^4 \sum_{i=1}^k \frac{1}{\epsilon^2} + \frac{r^2}{\epsilon^2} \leqslant \eta(k+1) \quad (29)$$

其中 η 为一常数. 同时由引理 2 可知

$$\cos\theta_k = -\frac{g_k^T d_k}{\|g_k\| \|d_k\|} \geqslant (2 - \sum_{j=0}^k \sigma_j^2) \frac{\|g_k\|}{\|d_k\|} \geqslant \frac{1-2\sigma_2}{1-\sigma_2} \frac{\|g_k\|}{\|d_k\|} \quad (30)$$

由 $\sigma_2 < \frac{1}{2}$ 和式(31) 可知常数 $\tau > 0$, 使得如下表达式成立

$$\sum_{k=0}^{\infty} \cos^2\theta_k \geqslant \left(\frac{1-2\sigma_2}{1-\sigma_2}\right) \sum_{k=0}^{\infty} \frac{\|g_k\|^2}{\|d_k\|^2} \geqslant \tau_2 \sum_{k=0}^{\infty} \frac{1}{k+1} \quad (31)$$

由于级数 $\sum_{k=0}^{\infty} \frac{1}{k+1}$ 是发散的, 可以得到 $\sum_{k=0}^{\infty} \cos^2\theta_k$ 也是发散的. 另外, 梯度函数 $g(x)$ 满足 Lipschitz 连续条件, 则有

$$\sigma_2 g_k^T d_k \leqslant g_{k+1}^T d_k = g_k^T d_k + (g_{k+1} - g_k)^T d_k \leqslant g_k^T d_k + \alpha_k L \|d_k\|^2 \quad (32)$$

从式(32) 中可以得出

$$\alpha_k \geqslant \frac{1-\sigma_2}{L \|d_k\|^2} g_k^T d_k \quad (33)$$

由强 Wolf 步长选择法则可得

$$f_{k+1} \leqslant f_k - \sigma_1 \frac{1-\sigma_2}{L} \|g_k\|^2 \cos^2\theta_k \quad (34)$$

对于以上证明过程可知, 当数列 $f(x_k)$ 是单调下降且下有界的, 则 $\|g_k\|^2 \cos^2\theta_k$ 是收敛的. 另外, 由

$\|g_k\| \geqslant \epsilon$, 可得 $\sum_{k=0}^{\infty} \cos^2\theta_k$ 是收敛的. 这样的证明结果与结论相互矛盾, 从而得到假设 $\|g_k\| \geqslant \epsilon$ 是不成立的, 本文算法的收敛性得证.

3 实验分析

上节中证明了本文的混合共轭梯度法的全局收敛性, 本节中结合共轭梯度法一些经典和改进的算法与我们提出的新改进混合共轭梯度法进行对比. 参数设置为 $\delta = 0.01$, $\sigma = 0.1$, 算法的终止条件为 $\|g_k\| \leqslant 1 \times 10^{-6}$, 或迭代次数大于 9 000. 表 1 是实验中采用的测试问题范例, 以表 1 中问题进行对比实验.

表 1 测试问题

编号	测试问题	编号	测试问题
1	Arwhead	2	Power
3	Quartc	4	Hager
5	Engvall	6	Diagonal
7	Liarwhd	8	Extend BD1
9	Extended Penalty	10	Extebde Matators
11	Generalized Quartic	12	Extended Tridiagonal
13	Cosine	14	Raydan1
15	Extended Denschnf	16	Nonscomp
17	Extended Denschnb	18	Almost Perturbed Quadratratic
19	Diagonal2	20	Extended Himmelblau
21	Extebde Tet	22	Generalized Tridiagonal1
23	Raydan2	24	Extend BD2

图 2 所示是本文方法与 PRP 共轭梯度算法在迭代时间方面的性能比较,从图 2 中分析可知,本文方法的迭代时间要少于传统的 PRP 共轭梯度法.

图 3 是本文方法和 HYBRID 方法在迭代时间性能上的对比,从实验结果分析可知,本文方法在迭代时间上要优于 HYBRID 方法. 在实际实验中,本文方法的运行时间比普通共轭梯度法减少了 10%左右. 因此,通过以上实验结果分析可知,本文提出的改进混合共轭梯度法要优于经典共轭梯度法.

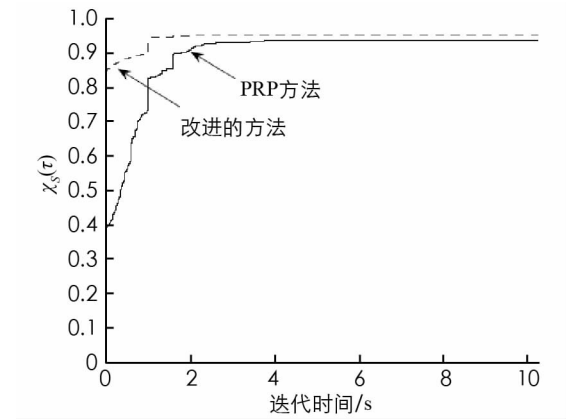


图 2 本文方法与 PRP 方法迭代时间性能对比

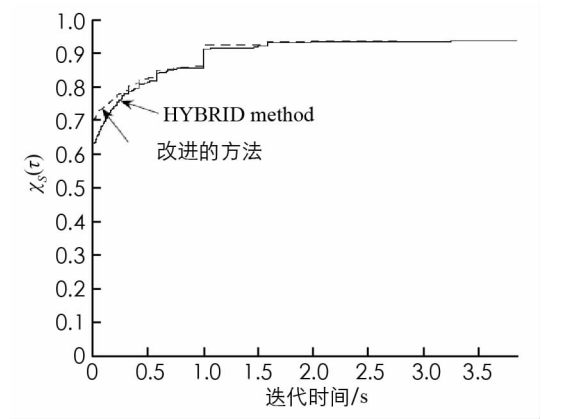


图 3 本文方法与 HYBRID 迭代时间性能对比

4 结 论

本文在已有共轭梯度算法求解无约束优化问题的基础上,提出了一种新混合共轭梯度方法,并给出了在强 wolfe 线搜索条件下的全局收敛性证明. 实验结果表明,本文方法在迭代时间性能方面要优于 PRP 方法和 HYBRID 方法.

参考文献:

[1] 王宜举,修乃华. 非线性最优化理论与方法 [M]. 北京: 科学出版社, 2012.

[2] 黎 勇, 韦增欣. 一种自动充分下降的共轭梯度法 [J]. 西南师范大学学报(自然科学版), 2016, 41(5): 36-40.

[3] 黄 海. 非线性无约束优化问题的新共轭梯度法 [J]. 河南大学学报(自然科学版), 2014, 44(2): 141-145.

[4] 焦宝聪, 陈兰平, 潘翠英. Goldstein 线搜索下混合共轭梯度法的全局收敛性 [J]. 计算数学, 2007, 29(2): 137-146.

[5] 刘 艳, 李 雷. 基于 PRP 共轭梯度的重构算法研究 [J]. 计算机技术与发展, 2016, 26(8): 55-59.

[6] 关 哲, 于宪伟. 标准 Wolfe 线搜索下修正的 DY 共轭梯度法 [J]. 西南师范大学学报(自然科学版), 2016, 41(1): 31-34.

[7] 周雪琴. 一个新的 PRP 共轭梯度法的全局收敛性 [J]. 重庆工商大学学报(自然科学版), 2015, 32(11): 31-33.

- [8] 何晓旭, 殷守林, 赵志刚. 一种新的无优化约束问题的混合 FR 和 PRP 共轭梯度算法 [J]. 沈阳师范大学学报(自然科学版), 2016, 34(1): 92-95.
- [9] 车彩丽. 基于 HS-DY 共轭梯度算法的概率布尔网络解法研究 [D]. 沈阳: 东北大学, 2014.
- [10] 刘 玲, 田志远, 商春雷. 带参数的混合共轭梯度算法及其收敛性研究 [J]. 青岛大学学报(自然科学版), 2016, 29(3): 16-20, 24.

A New Mixed Conjugate Gradient Method for Solving Unconstrained Optimization Problems

TANG Tian-guo

Department of Electronic Information Engineering, Nanchong Vocational and Technical College, Nanchong Sichuan 637000, China

Abstract: Based on existing conjugate gradient method, a new conjugate gradient method is proposed to solve the unconstrained optimization problem. An approximate method has been used to approximate the Hessian matrix and to overcome the problem of large computational complexity in the Hessian matrix of traditional Newton method. And under strong Wolfe line search, the global convergence of the conjugate gradient algorithm is proved. The experimental results show that, compared with the PRP(Polak-Ribiere-Polyak) method and HYBRID method based on BFGS(Broyden-Fletcher-Goldfarb-Shanno), the iterative time of the hybrid conjugate gradient algorithm proposed in this paper is less than that of former two methods, which shows that the method is feasible and effective.

Key words: conjugate gradient method; unconstrained optimization; strong Wolfe line search; global convergence

责任编辑 夏 娟