

精确线搜索下一种新的混合共轭梯度法

景书杰, 王慧婷, 牛海峰, 陈 耀

(河南理工大学数学与信息科学学院, 河南 焦作 454000)

摘 要: 本文对于大规模无约束优化问题提出了一种新的混合 β_k 公式, 从而提出了一种具有充分下降性的混合共轭梯度法. 利用精确线搜索步长规则, 在适当的假设下证明了新算法的全局收敛性.

关键词: 无约束最优化; 共轭梯度法; 精确线搜索; 全局收敛性

MR(2010) 主题分类号: 90C30; 65K10

中图分类号: O221.2

文献标识码: A

文章编号: 0255-7797(2018)03-0520-05

1 引言

考虑如下无约束优化问题

$$\min_{x \in R^n} f(x), \quad (1.1)$$

其中 $f(x)$ 是 $R^n \rightarrow R$ 上的连续可微函数, 用非线性共轭梯度法求解无约束优化问题 (1.1), 点列 $\{x_k\}$ 的迭代格式为

$$x_{k+1} = x_k + \alpha_k d_k, \quad (1.2)$$

这里的 α_k 为步长. 本文选用精确线搜索计算 α_k , 即每步迭代中选择 α_k 满足

$$\alpha_k = \min\{\alpha | \nabla f(x_k + \alpha d_k)^T d_k = 0, \alpha > 0\}. \quad (1.3)$$

选取步长因子 α_k 最好的方法就是使目标函数沿着搜索方向 d_k 达到极小, 从理论上来说, 精确线搜索所得到的步长因子有着最好的下降量. 例如 Zoutendijk^[1] 证明了采取精确线搜索的 FR 方法对一般非凸函数总收敛; 从文献 [2] 中的结果可知用精确线搜索的 PRP 方法对一致凸函数全局收敛; Rivaie^[3] 提出了一个新算法在精确线搜索下有更好的结果.

搜索方向 d_k 的迭代格式为

$$d_k = \begin{cases} g_k, & k = 1, \\ g_k + \beta_k d_{k-1}, & k \geq 2, \end{cases} \quad (1.4)$$

其中 $g_k = \nabla f(x_k)$, 如果用 θ_k 表示向量 d_k 与 $-g_k$ 的夹角, 则有

$$\cos \theta_k = \frac{-g_k^T d_k}{\|g_k\| \|d_k\|}, \quad (1.5)$$

*收稿日期: 2016-11-02

接收日期: 2017-04-01

基金项目: 国家自然科学基金资助项目 (U1504104); 河南省教育厅教育厅基础研究计划资助项目 (15A110024).

作者简介: 景书杰 (1965-), 男, 河南长垣, 教授, 主要研究方向: 最优化理论与应用.

这里的 β_k 为一标量, 著名的 β_k 公式有 (可参看文献 [4-8])

$$\begin{aligned}\beta_k^{\text{FR}} &= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_k^{\text{PRP}} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}, \\ \beta_k^{\text{HS}} &= \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}, \quad \beta_k^{\text{DY}} = \frac{\|g_k\|^2}{d_{k-1}^T(g_k - g_{k-1})},\end{aligned}$$

其中 $\|\cdot\|$ 表示欧式范数. 通常情况下, FR 和 DY 方法有很好的收敛性, 而 PRP 和 HS 方法却有很好的数值效果. 学者们为了寻找既能保证收敛性又可以有良好数值效果的算法, 在以上公式的基础上, 一方面对 β_k 进行改进例如文献 [3, 9, 10], 另一方面将不同的 β_k 公式进行混合 [11-13]. 最近, 文献 [3] 中给出了一个新的参数公式

$$\beta_k^{\text{RMIL}} = \frac{g_k^T(g_k - g_{k-1})}{\|d_{k-1}\|^2}, \quad (1.6)$$

并得到了该算法在精确线搜索下的全局收敛性. 受文献 [12] 的启发, 取 β_k^{New} 为

$$\beta_k^{\text{New}} = \begin{cases} 0, & \|g_k\|^2 < |g_k^T g_{k-1}|, \\ \mu \frac{\|g_k\|^2 - |g_k^T g_{k-1}|}{|g_k^T d_{k-1}| + \|d_{k-1}\|^2}, & \|g_k\|^2 \geq |g_k^T g_{k-1}|, \end{cases} \quad (1.7)$$

其中 μ 为参数且 $0 < \mu \leq 1$.

2 算法及其性质

本文讨论一种新的混合共轭梯度法, 其中

$$\beta_k = \begin{cases} \beta_k^{\text{New}}, & \|g_k\|^2 \geq |g_k^T g_{k-1}|, \\ \beta_k^{\text{RMIL}}, & \text{其他}. \end{cases} \quad (2.1)$$

算法 A 步骤 1 给定 $\varepsilon > 0, x_1 \in R^n, 0 < \mu \leq 1, d_1 = -g_1, k := 1$.

步骤 2 若 $\|g_k\| < \varepsilon$, 则停止; 否则转步骤 3.

步骤 3 由精确线搜索计算步长 α_k , 使其满足 (1.3) 式.

步骤 4 令 $x_{k+1} = x_k + \alpha_k d_k$, 求 g_{k+1} , 并用 (2.1) 式试求 β_{k+1} .

步骤 5 令 $d_{k+1} = -g_{k+1} + \beta_{k+1} d_k$, 令 $k = k + 1$; 转步骤 2.

引理 2.1 对任意 $k \geq 1$, 算法 A 产生的搜索方向 d_k 满足 $g_k^T d_k \leq -C\|g_k\|^2$, 其中 $C \geq 0$.

证 当 $k = 1$ 时, $g_1^T d_1 = -\|g_1\|^2 < 0$, 故结论成立.

(i) 当 $\beta_k = \beta_k^{\text{New}}$ 时, 若 $\beta_k^{\text{New}} = 0$, 有 $g_k^T d_k = -\|g_k\|^2$, 结论成立. 否则有

$$\beta_k^{\text{New}} = \mu \frac{\|g_k\|^2 - |g_k^T g_{k-1}|}{|g_k^T d_{k-1}| + \|d_{k-1}\|^2} \leq \mu \frac{\|g_k\|^2}{|g_k^T d_{k-1}| + \|d_{k-1}\|^2} \leq \mu \frac{\|g_k\|^2}{|g_k^T d_{k-1}|}.$$

因此当 $k > 1$ 时,

$$g_k^T d_k = -\|g_k\|^2 + \beta_k^{\text{New}} g_k^T d_{k-1} \leq -\|g_k\|^2 + \mu \frac{\|g_k\|^2}{|g_k^T d_{k-1}|} g_k^T d_{k-1} = -(1 - \mu)\|g_k\|^2. \quad (2.2)$$

因为 $0 < \mu \leq 1$, 故 $g_k^T d_k \leq -C\|g_k\|^2$, 其中 $C \geq 0$.

(ii) 当 $\beta_k = \beta_k^{\text{RMIL}}$, $k > 1$ 时, 对式 (1.3) 有

$$g_k^T d_k = -\|g_k\|^2 + \beta_k^{\text{RMIL}} g_k^T d_{k-1}.$$

对于精确线搜索, 有 $g_k^T d_{k-1} = 0$. 所以 $g_k^T d_k = -\|g_k\|^2$, 故 $g_k^T d_k \leq -C\|g_k\|^2$ 成立, 证毕.

引理 2.2 同引理 2.1 中的条件, 由式 (1.5) 和 (2.1) 可得 $|\beta_k| \leq \beta_k^{\text{RMIL}}$.

证 因为

$$\beta_k^{\text{RMIL}} = \frac{g_k^T (g_k - g_{k-1})}{\|d_{k-1}\|^2} = \frac{\|g_k\|^2 - |g_k^T g_{k-1}|}{\|d_{k-1}\|^2},$$

由 (2.1) 式知, 当 $\beta_k = \beta_k^{\text{New}}$ 时, $\|g_k\|^2 \geq |g_k^T g_{k-1}|$, 所以

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

所以 $-\beta_k^{\text{RMIL}} \leq \beta_k^{\text{New}} \leq \beta_k^{\text{RMIL}}$, 故有 $|\beta_k| \leq \beta_k^{\text{RMIL}}$. 证毕.

3 全局收敛性

假设

(H1) 目标函数 $f(x)$ 在水平集 $L_0 = \{x \in R^n | f(x) \leq f(x_1)\}$ 上有下界, 其中 x_1 为初始点.

(H2) 目标函数 $f(x)$ 在水平集 L_0 的一个邻域 N 内连续可微, 且梯度函数 $g(x)$ 满足 Lipschitz 连续, 即存在常数 $L > 0$, 使

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in N. \quad (3.1)$$

引理 3.1 若 (H1), (H2) 成立, 考虑一般方法 $x_{k+1} = x_k + \alpha_k d_k$, 其中 d_k 满足 $d_k^T g_k < 0$, 步长 α_k 满足精确线搜索 (1.6) 式, 则有

$$\sum_{k \geq 1} \frac{(d_k^T g_k)^2}{\|d_k\|^2} < +\infty. \quad (3.2)$$

证 由 (1.6) 式, $\alpha_k = \min\{\alpha | \nabla f(x_k + \alpha d_k)^T d_k = 0, \alpha > 0\}$, 即 $g(x_k + \alpha_k d_k)^T d_k = 0$. 再由 Lipschitz 条件 (3.1), 有 $\|g(x_k + \alpha_k d_k) - g(x_k)\| \cdot \|d_k\| \leq L\alpha_k \|d_k\|^2$. 由 Cauchy-Schwartz 不等式可得

$$\|g(x_k + \alpha_k d_k) - g(x_k)\| \cdot \|d_k\| \geq [g(x_k + \alpha_k d_k) - g_k]^T d_k = -g_k^T d_k.$$

所以 $L\alpha_k \|d_k\|^2 \geq -g_k^T d_k$, 即 $\alpha_k \geq -\frac{g_k^T d_k}{L\|d_k\|^2}$. 又因为由假设可知, 对一切 $\alpha > 0$ 都成立

$$f_{k+1} \leq f_k + \alpha d_k^T g_k + \frac{\alpha^2}{2} L \|d_k\|^2.$$

令 $\alpha^* = \frac{g_k^T d_k}{L\|d_k\|^2}$, 由精确线搜索准则可知

$$\begin{aligned} f_k - f_{k+1} &\geq f_k - f(x_k + \alpha^* d_k) \geq \alpha^* d_k^T g_k - \frac{\alpha^{*2}}{2} L \|d_k\|^2 \\ &= \frac{(g_k^T d_k)^2}{L\|d_k\|^2} - \frac{(g_k^T d_k)^2}{2L\|d_k\|^2} = \frac{(g_k^T d_k)^2}{2L\|d_k\|^2} = C \frac{g_k^T d_k}{\|d_k\|^2}, \end{aligned}$$

其中 $C = \frac{1}{2L}$, 即 $f_k - f_{k+1} \geq C \frac{g_k^T d_k}{\|d_k\|^2}$, 对上式从 $k = 1, 2, \dots$, 求和, 因为 $f(x)$ 有下界, 故结论成立.

定理 3.1 设目标函数满足假设 (H1), (H2), 若存在常数 $m > 0$, 使得 $\|g(x)\| \leq m$, $\forall x \in L_0$, 迭代点列 $\{x_k\}$ 由算法 A 产生, 则有

$$\lim_{k \rightarrow \infty} \inf \|g_k\| = 0. \quad (3.3)$$

证 假设定理不成立, 所以存在一个常数 $C > 0$, 有 $\|g(x)\| \leq C$. 由 $d_k + g_k = \beta_k d_{k-1}$, 对等式两端取模平方, 并移项得到

$$\|d_k\|^2 = (\beta_k)^2 \|d_{k-1}\|^2 - 2g_k^T d_k - \|g_k\|^2,$$

由引理 2.2 可知 $|\beta_k| \leq \beta_k^{\text{RMIL}}$, 所以

$$\|d_k\|^2 = (\beta_k^{\text{RMIL}})^2 \|d_{k-1}\|^2 - 2g_k^T d_k - \|g_k\|^2.$$

两边除以 $(g_k^T d_k)^2$ 得

$$\begin{aligned} \frac{\|d_k\|^2}{(g_k^T d_k)^2} &\leq \frac{(\beta_k^{\text{RMIL}})^2 \|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{\|g_k\|^2}{(g_k^T d_k)^2} - \frac{2}{g_k^T d_k} \\ &= \frac{(\beta_k^{\text{RMIL}})^2 \|d_{k-1}\|^2}{(g_k^T d_k)^2} - \left(\frac{1}{\|g_k\|} + \frac{\|g_k\|}{g_k^T d_k}\right)^2 + \frac{1}{\|g_k\|^2} \\ &\leq \frac{(\beta_k^{\text{RMIL}})^2 \|d_{k-1}\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_k\|^2}. \end{aligned}$$

又因为

$$\beta_k^{\text{RMIL}} = \frac{g_k^T (g_k - g_{k-1})}{\|d_{k-1}\|^2} = \frac{\|g_k\|^2 - g_k^T g_{k-1}}{\|d_{k-1}\|^2} \leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2},$$

所以

$$\begin{aligned} \frac{\|d_k\|^2}{(g_k^T d_k)^2} &\leq \left(\frac{\|g_k\|^2}{\|d_{k-1}\|^2}\right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_k\|^2} = \frac{\|g_k\|^2}{\|d_{k-1}\|^2 \|d_k\|^2} + \frac{1}{\|g_k\|^2} \\ &\leq \frac{1}{\|d_{k-1}\|^2} + \frac{1}{\|g_k\|^2} \leq \dots \leq \sum_{i=1}^k \frac{1}{\|g_i\|^2} \leq \frac{k}{c^2}. \end{aligned}$$

所以

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{c^2}{k},$$

即

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = +\infty.$$

与引理 3.1 中的 (3.2) 式矛盾, 故

$$\lim_{k \rightarrow \infty} \inf \|g_k\| = 0. \quad (3.4)$$

证毕.

参 考 文 献

- [1] Zoutendijk G. Nonlinear programming, computational methods [J]. Integ. Nonl. Prog., 1970: 37–86.
- [2] Powell M J D. Restart procedures for the conjugate gradient method [J]. Math. Prog., 1977, 12(1): 241–254.
- [3] Rivaie M, Mamat M, June L W, Mohd I. A new class of nonlinear conjugate gradient coefficients with global convergence properties [J]. Appl. Math. Comp., 2012, 218: 11323–11332.
- [4] Hestenes M R, Stiefel E L. Methods of conjugate gradients for solving linear systems [J]. J. Res. Nat. Bureau Stand., 1952, 5(49): 409–436.
- [5] Fletcher R, Reeves C. Functions minimization by conjugate gradients [J]. Comp. J., 1964, 7(2): 149–154.
- [6] Polak E, Ribière G. Note sur la convergence de méthodes de directions conjuguées [J]. Rev. Fran. Inform. Rech. Opérationelle, 1969, 16(3): 35–43.
- [7] Polyak B T. The conjugate gradient method in extreme problems [J]. USSR Comp. Math. Math. Phys., 1969, 9: 94–112.
- [8] Dai Y H, Yuan Y X. A Nonlinear conjugate gradient method with a strong global convergence property [J]. SIAM J. Optim., 1999, 10: 177–182.
- [9] Rivaie M, Mamat M, Abashar A. A new class of nonlinear conjugate gradient coefficients with exact and inexact line searches [J]. Appl. Math. Comp., 2015, 268: 1152–1163.
- [10] Xu Z S. A class of new conjugate gradient methods [J]. J. Math., 2002, 22(1): 27–30.
- [11] 江羨珍, 韩麟, 简金宝. Wolfe 线搜索下一个全局收敛的混合共轭梯度法 [J]. 计算数学, 2012, 34(1): 103–112.
- [12] 戴志峰, 陈兰平. 一种混合的 HS-DY 共轭梯度法 [J]. 计算数学, 2005, 27(4): 429–436.
- [13] 景书杰, 邓涛. 精确线搜索下具有充分下降性的混合共轭梯度法 [J]. 河南理工大学学报 (自然科学版), 2010, 29(2): 266–273.

A NEW CLASS OF MIXED CONJUGATE GRADIENT METHOD
WITH EXACT LINE SEARCH

JING Shu-jie, WANG Hui-ting, NIU Hai-feng, CHEN Yao

*(School of Mathematics and Information Science, Henan Polytechnic University,
Jiaozuo 454000, China)*

Abstract: In this paper, a new mixed iterative formula of coefficient β_k is proposed for large-scale unconstrained optimization problems, and a mixed conjugate gradient method with sufficient descent is proposed. By using the exact line search step rules, we prove the global convergence of the new algorithm under the appropriate assumptions.

Keywords: unconstrained optimization; conjugate gradient; exact line search; global convergence

2010 MR Subject Classification: 90C30; 65K10