DOI:10.13656/j.cnki.gxkx.2005.03.003
广西科学 Guangxi Sciences 2005, 12(3):167~169

广义 Wolfe 线搜索下共轭梯度法的全局收敛性 *

Global Convergence of Conjugate Gradient Method with General Wolfe Line Search

韦增欣, 武小平, 刘利英

Wei Zengxin, Wu Xiaoping, Liu Liying

(广西大学数学与信息科学学院, 广西南宁 530004)

(Coll. of Math. & Info. Sci., Guangxi Univ., Nanning, Guangxi, 530004, China)

摘要: 给出一类求解非线性无约束优化问题修正的共轭梯度类型公式和算法, 并证明该公式在广义 Wolfe 线搜索下具有充分下降性和全局收敛性.

关键词: 共轭梯度法 Wolfe 线搜索 充分下降性 全局收敛性

中图法分类号: O224 文献标识码: A 文章编号: 1005-9164(2005)03-0167-03

Abstract: A new kind of conjugate gradient method and algorithm for solving nonlinear unconstrainted optimization problem was proposed. The sufficient descent property and global convergence of the new conjugate gradient method with general Wolfe line search were proved.

Key words: conjugate gradient, Wolfe line search, sufficient descent, global convergence

本文主要考虑无约束优化问题

$$\min\{f(x) \mid x \in \mathscr{R}^n\}, \tag{0.1}$$

其中 $f: \mathscr{R}^n \to \mathscr{R}$ 为一阶可微的非线性函数, 其梯度记为 $g: \mathscr{R}^n \to \mathscr{R}^n$. 共轭梯度法是用来求解无约束优化问题(0.1)的一种方法, 其迭代公式为:

$$x_{k+1} = x_k + t_k d_k, \tag{0.2}$$

其中 t_k 为搜索步长, 一般采用某个线搜索来获得, 常用的步长规则主要有:

(1) Amijo 规则. 寻找一个 $t_k = \beta^{m_k} s$, 其中 $\beta \in (0,1)$, $s > 0$, m_k 是最小的正整数 m, 并满足:

$$f(x_k + t_k d_k) - f(x_k) \leqslant \delta \beta^m s g_k^T d_k,\ \delta \in (0,1). \tag{0.3}$$

(2) Amijo-Goldstein 规则. 寻找一个 $t_k > 0$ 并满足:

$$f(x_k + t_k d_k) - f(x_k) \leqslant \delta t_k g_k^T d_k, \tag{0.4}$$

$$f(x_k + t_k d_k) - f(x_k) \geqslant (1-\delta) t_k g_k^T d_k,\quad \delta \in (0, \frac{1}{2}). \tag{0.5}$$

(3) 弱 Wolfe-Powell 规则. 寻找一个 $t_k > 0$ 并满足(0.4)和

$$g_{k+1}^T d_k \geqslant \sigma t_k g_k^T d_k,\ \sigma \in (\delta, 1). \tag{0.6}$$

(4) 强 Wolfe-Powell 规则. 寻找一个 $t_k > 0$, 并满足(0.4)和

$$|g_{k+1}^T d_k| \leqslant \sigma |g_k^T d_k|,\ \sigma \in (\delta, 1). \tag{0.7}$$

(5) 广义 Wolfe 规则. 寻找一个 $t_k > 0$, 并满足:

$$f(x_k + t_k d_k) - f(x_k) \leqslant \delta t_k g_k^T d_k, \tag{0.8}$$

$$\sigma_1 g_k^T d_k \leqslant g_{k+1}^T d_k \leqslant -\sigma_2 g_k^T d_k, \tag{0.9}$$

其中, $\delta, \sigma_1, \sigma_2 \in (0,1)$.

在(0.2)式中 d_k 为搜索方向, 被定义为:

$$d_k = \begin{cases} -g_k, & k = 1, \\ -g_k + \beta_k d_{k-1}, & k \geqslant 2. \end{cases} \tag{0.10}$$

其中, β_k 是标量, 文献[1~7]已给出选取 β_k 的几种公式.

选取 β_k 的几种公式在某些线搜索下的全局收敛性已被广泛研究, 其中 Powell 证明了 FR 方法在精确线搜索下具有全局收敛性[8]; Al-Baali 在非精确线搜索下证明了 FR 方法具有全局收敛性但数值结果较差[9]; Polak-Ribieer 方法有很好的数值结果但其局收敛性较差, 因此 Touati-Ahmed 和 Story 把 β_k^{FR}, β_k^{PR} 结合起来, 得到了具有较好的数值结果和全局收敛性的方法[10]; Han Ji-Ye 等[11] 把公式中 β_k 的取值范围进一步扩大, 并证明了在强 Wolfe 线搜索下的全局收敛性. 本文吸收文献[10,12]的思想得出一类新的共轭梯度类型公式和算法, 并证明该公式在广义 Wolfe 线搜索下具有充分下降性和全局收敛性.

收稿日期: 2004-11-23
修回日期: 2005-01-12
作者简介: 韦增欣(1962-), 男, 广西武鸣人, 教授, 博士, 主要从事优化管理研究.
*国家自然科学基金(10161002)和广西自然科学基金(0135004)资助项目.

1 新的共轭梯度类型公式及其充分下降性

在研究的过程中首先要保证选取的步长 t_k 和搜索方向 d_k 能使相应的共轭梯度算法是下降的. 即满足

$$g_k^T d_k < 0, \quad (1.1)$$

而且更希望构造的 β_k 满足充分下降条件:

$$g_k^T d_k \leqslant -c \| g_k \|^2, c > 0, \quad (1.2)$$

因为对共轭梯度法来说, 充分下降性是一个非常重要的性质, 它对于保证算法的全局收敛性有很好的作用[13]. 受到文献[10,12]的启发作者构造了一类新公式:

$$\sigma \left| \frac{\beta_k}{\beta_k^{CD}} \right| \leqslant \hat{\sigma}, \quad (1.3)$$

其中 $\hat{\sigma} \in (0,1)$, $\sigma = \max\{\sigma_1, \sigma_2\}$, σ_1, σ_2 满足(0.9)式. 该公式在广义 Wolfe 线搜索下具有充分下降性和全局收敛性. 下面给出新公式充分下降性质的证明.

定理 1 在广义 Wolfe 线搜索条件(0.8)式和(0.9)式下, 若 β_k 满足(1.3)式, 则 d_k 具有充分下降性, 即满足下降条件(1.2)式.

证明 先证明:

$$1 - \hat{\sigma} \leqslant -\frac{g_k^T d_k}{g_k^T g_k}. \quad (1.4)$$

当 $k = 1$ 时, 上式显然成立.

当 $k \geqslant 2$ 时,

$$-\frac{g_k^T d_k}{g_k^T g_k} = -\frac{g_k^T(-g_k + \beta_k d_{k-1})}{g_k^T g_k} = 1 - \beta_k \frac{g_k^T d_{k-1}}{g_k^T g_k} \geqslant 1 - |\beta_k| \frac{|g_k^T d_{k-1}|}{g_k^T g_k} \geqslant 1 + |\beta_k| \sigma \frac{g_{k-1}^T d_{k-1}}{g_k^T g_k} \geqslant 1 + \hat{\sigma}\left(-\frac{g_k^T g_k}{g_{k-1}^T d_{k-1}}\right) \frac{g_{k-1}^T d_{k-1}}{g_k^T g_k} \geqslant 1 - \hat{\sigma},$$

即(1.4)式成立.

从而有:

$$g_k^T d_k \leqslant -(1 - \hat{\sigma}) \| g_k \|^2,$$

即 $g_k^T d_k \leqslant -c \| g_k \|^2$,

其中 $c = 1 - \hat{\sigma}$. 由此可见新公式具有充分下降性.

2 新的共轭梯度算法及其全局收敛性

求解无约束优化问题共轭梯度法的新算法如下.

步骤 1 任意给定 $x_1 \in R^n$, 计算 g_1, 若 $g_1 = 0$, 则停; 否则, 转步骤 2.

步骤 2 令 $d_1 = -g_1$, $k = 1$.

步骤 3 令 $x_{k+1} = x_k + t_k d_k$, t_k 满足(0.8)和(0.9)式.

步骤 4 计算 g_{k+1}, 若 $\| g_{k+1} \| = 0$, 则停; 否则, 令 $k := k + 1$, 转步骤 5.

步骤 5 令 $d_k = -g_k + \beta_k d_{k-1}$, 其中 β_k 满足(1.3)式及 $\sum_{j=1}^{k} (\prod_{i=j+1}^{k} \beta_i)^2 \| g_j \|^4 \leqslant dk$, $k \in N$, $\prod_{i=k+1}^{k} \beta_i^2 = 1$.

为了证明该方法的全局收敛性, 给出下面的假设和引理.

假设 1 水平集 $\Omega = \{x \in \mathscr{R}^n : f(x) \leqslant f(x_1)\}$ 有界.

假设 2 在水平集上梯度 g Lipschitz 连续, 即对于任意 $x, y \in \Omega$, $\| g(x) - g(y) \| \leqslant M \| x - y \|$ 成立, M 为任意常数.

引理 1 若(0.8), (0.9)式和假设 2 成立, 则对于任意 $k \in N$, Zoutendijk 条件成立, 即

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\| d_k \|^2} < +\infty. \quad (2.1)$$

证明 由(0.9)式得:

$$(\sigma_1 - 1) g_k^T d_k \leqslant (g_{k+1} - g_k)^T d_k,$$

$$(1 - \sigma_1) | g_k^T d_k | \leqslant \| (g_{k+1} - g_k \| \| d_k \| \leqslant M t_k \| d_k \|^2,$$

$$t_k \geqslant \frac{(1 - \sigma_1) | g_k^T d_k |}{M \| d_k \|^2}.$$

由(0.8)式得 $f(x_k) - f(x_{k+1}) \geqslant -\delta t_k g_k^T d_k$, 将其代入上式,

$$f(x_k) - f(x_{k+1}) \geqslant \delta \frac{1 - \sigma_1}{M} \frac{(g_k^T d_k)^2}{\| d_k \|^2},$$

$$\sum_{k=1}^{\infty} f(x_k) - f(x_{k+1}) \geqslant \sum_{k=1}^{\infty} \delta \frac{1 - \sigma_1}{M} \frac{(g_k^T d_k)^2}{\| d_k \|^2},$$

两边取极限得:

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

引理 2 若假设 1, 假设 2 成立, 则在广义 Wolfe 线搜索下有:

$$\sum_{k=1}^{\infty} (\gamma_k + \gamma_{k-1})^2 < +\infty, \gamma_k = -\frac{g_k^T d_k}{\| d_k \|}. \quad (2.2)$$

证明 由引理 1 知

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

$$\sum_{k=1}^{\infty} \gamma_k^2 < +\infty,$$

$$\sum_{k=1}^{\infty} (\gamma_k + \gamma_{k-1})^2 = \sum_{k=1}^{\infty} (\gamma_k^2 + \gamma_{k-1}^2 + 2\gamma_k \gamma_{k-1}) \leqslant 2\sum_{k=1}^{\infty} (\gamma_k^2 + \gamma_{k-1}^2) < +\infty.$$

引理 3 若 $\| g_k \| \geqslant c_2$, β_k 满足(1.3)式, $\sum_{j=1}^{k} (\prod_{i=j+1}^{k} \beta_i)^2 \| g_j \|^4 \leqslant dk$, $k \in N$, $\prod_{i=k+1}^{k} \beta_i^2 = 1$, 则

$$\| d_k \| \leqslant \frac{(1+2\hat{\sigma})^{\frac{1}{2}}}{c_2} c^{\frac{1}{2}} k^{\frac{1}{2}}. \quad (2.3)$$

证明 由 $\sigma = \max\{\sigma_1, \sigma_2\}$ 及(0.9)式,(0.10)式,(1.3)式知

$\| d_k \|^2 = \| g_k \|^2 - 2\beta_k g_k^T d_{k-1} + \beta_k^2 \| d_{k-1} \|^2 \leqslant \| g_k \|^2 + 2 \left| \frac{\beta_k}{\beta_k^{CD}} \right| |\beta_k^{CD}| |g_k^T d_{k-1}| + \beta_k^2 \| d_{k-1} \|^2 \leqslant \| g_k \|^2 + 2\hat{\sigma} \| g_k \|^2 + \beta_k^2 \| d_{k-1} \|^2 \leqslant (1+2\hat{\sigma}) \| g_k \|^2 + \beta_k^2 \| d_{k-1} \|^2 \leqslant (1+2\hat{\sigma}) \| g_k \|^2 + (1+2\hat{\sigma}) \| g_{k-1} \|^2 \beta_k^2 + \beta_k^2 \beta_{k-1}^2 \| d_{k-2} \|^2 \leqslant \cdots \leqslant \frac{1+2\hat{\sigma}}{c_2^2} \sum_{j=1}^{k} (\prod_{i=j+1}^{k} \beta_i)^2 \| g_j \|^4 \leqslant \frac{1+2\hat{\sigma}}{c_2^2} ck,$

因此(2.3)式成立,证毕.

定理2 如果假设1,假设2成立,β_k 满足公式(1.3),t_k 由广义Wolfe线搜索(0.8)式和(0.9)式产生,则该算法具有全局收敛性,即

$$\lim_{k \to \infty} \inf \| g_k \| = 0. \quad (2.4)$$

证明 由共轭梯度法迭代公式(0.2)式及(0.9)式有:

$\| d_k \|^2 = \| g_k \|^2 - 2\beta_k g_k^T d_{k-1} + \beta_k^2 \| d_{k-1} \|^2 \geqslant \| g_k \|^2 - 2\beta_k g_k^T d_{k-1} \geqslant \| g_k \|^2 + 2\sigma_2 \beta_k g_{k-1}^T d_{k-1}$,

由线搜索知 $g_{k-1}^T d_{k-1} < 0$, $|\beta_k| \geqslant \beta_k$,

$\| g_k \|^2 + 2\sigma_2 \beta_k g_{k-1}^T d_{k-1} \geqslant \| g_k \|^2 + 2\sigma_2 |\beta_k| (g_{k-1}^T d_{k-1}) \geqslant \| g_k \|^2 - 2\sigma_2 |\beta_k| \| d_{k-1} \| \gamma_{k-1} \geqslant \| g_k \|^2 - 2\sigma_2 \gamma_{k-1} (\| d_k \| + \| g_k \|)$,

由假设1得,$c_3 = \sup \| g_k \| < +\infty$,则

$\| g_k \|^2 \leqslant \| d_k \|^2 + 2\sigma_2 \gamma_{k-1} (\| d_k \| + \| g_k \|) \leqslant \| d_k \|^2 + 2\sigma_2 \gamma_{k-1} (\| d_k \| + c_3)$.

若 $\lim_{k \to \infty} \inf \| d_k \| = 0$,则 $\lim_{k \to \infty} \inf \| g_k \| = 0$;若 $\lim_{k \to \infty} \inf \| d_k \| > 0$,则存在 $c_4 > 0$ 使 $\| d_k \| \geqslant c_4$;若 $\lim_{k \to \infty} \inf \| g_k \| = 0$ 不成立,设 $\| g_k \| \geqslant c_2$ 则由引理3有

$\frac{1}{\| d_k \|} \geqslant \frac{1}{ak^{\frac{1}{2}}}$,其中 $a = \| d_k \| \leqslant \frac{((1+2\hat{\sigma})c)^{\frac{1}{2}}}{c_2}$,则

$$\sum_{k=2}^{\infty} \frac{1}{\| d_k \|^2} \geqslant \sum_{k=2}^{\infty} \frac{1}{a^2 k} = +\infty. \quad (2.5)$$

由 $\| d_k \| \geqslant |\beta_k| \| d_{k-1} \| - \| g_k \|$ 得

$$\frac{|\beta_k| \| d_{k-1} \|}{\| d_k \|} \leqslant 1 + \frac{\| g_k \|}{\| d_k \|}, \quad (2.6)$$

又由 $-g_k^T d_k = \| g_k \|^2 - \beta_k g_k^T d_{k-1}$, $\| g_k \| \geqslant c_2$ 及(0.9)式和(2.6)式得:

$\gamma_k = -\frac{g_k^T d_k}{\| d_k \|} = \frac{\| g_k \|^2}{\| d_k \|} - \frac{\beta_k g_k^T d_{k-1}}{\| d_k \|} \geqslant \frac{c_2^2}{\| d_k \|} - \frac{\sigma_2 |\beta_k| g_{k-1}^T d_{k-1}}{\| d_k \|} \geqslant \frac{c_2^2}{\| d_k \|} - \frac{\sigma_2 |\beta_k|}{\| d_k \|} \| d_{k-1} \| \gamma_{k-1} \geqslant \frac{c_2^2}{\| d_k \|^2} - \sigma_2 (1 + \frac{\| g_k \|}{\| d_k \|}) \gamma_{k-1}$.

由上式及 $c_3 = \sup \| g_k \| < \infty$, $\| d_k \| \geqslant c_4$ 有:

$\frac{1}{\| d_k \|} \leqslant \frac{\gamma_k}{c_2^2} + \frac{\sigma_2}{c_2^2} (1 + \frac{\| g_k \|}{\| d_k \|}) \gamma_{k-1} \leqslant \frac{\gamma_k}{c_2^2} + \frac{\sigma_2}{c_2^2} (1 + \frac{c_3}{c_4}) \gamma_{k-1} \leqslant \max\{\frac{1}{c_2^2}, \frac{\sigma_2}{c_2^2} (1 + \frac{c_3}{c_4})\} (\gamma_k + \gamma_{k-1})$,

从而:

$\sum_{k=2}^{\infty} \frac{1}{\| d_k \|^2} \leqslant (\max\{\frac{1}{c_2^2}, \frac{\sigma_2}{c_2^2} (1 + \frac{c_3}{c_4})\})^2 \sum_{k=2}^{\infty} (\gamma_k + \gamma_{k-1})^2 < +\infty$.

该式与(2.5)式矛盾,所以 $\lim_{k \to \infty} \inf \| g_k \| = 0$. 定理证毕.

参考文献:

[1] Hestenes M R, Stiefel E. Method of conjugate gradient for solving linear equations[J]. J Res Nat Bur Stand, 1952, 49: 409-436.

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

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

[4] Polyak B T. The conjugate gradient methodin extreme problems[J]. USSR Comp Math and Math Phys, 1969, 9: 94-112.

[5] Fletcher R. Practical method of optimization[M]. Vol I: unconstrained optimization. 2nd edition. New York: Wiley, 1987.

[6] Liu Y, Storey C. Efficient generalized conjugate gradient algorithms, part 1: theory[J]. Journal of Optimization Theory and Application, 1992, 69: 129-137.

[7] Dai Y, Yuan Y. A nonlinear conjugate gradientwith a strong global convergence properties[J]. SIAM Journal of Optimization, 1999, 10: 177-182.

[8] Powell M J D. Nonconvex minimization calculations and the conjugate gradient methd[A]. Lecture Notes in Mathematics: Vol 1066[C]. Belin: Springer-Verlag, 1984. 122-141.

[9] Al-Baali M. Descent property and global convergence of the Fletcher-Reeves method with inexact line search[J]. IMA J Numer Anal, 1985, (5): 121-124.

[10] Touati-Ahamed D, Storey C. Globally convergence hybrid conjugate gradient methods[J]. Joural of Optimization Theory and Applications, 1990, 64(2): 379-397.

[11] Han Jiye, Liu Guanghui, Yin Hongxia. Convergence properties of conjugate gradient methods with strong Wolfe line search[J]. Systems Science and Mathematics Science, 1998. 11(2): 112-116.

[12] Du Shouqiang, Chen Yuanyuan, Qi Longxin. Convergence properties of conjugate gradient method with general Wolfe line search[J]. Journal of Qufu Normal University, 2003, 29(3): 7.

[13] Dai Y. Further insight into the convergence of the Fletcher-Reeves method[J]. Sciences in China (Series A), 1999, 42(9): 905-916.

(责任编辑:韦廷宗 邓大玉)