第 1 卷 中国农业大学学报 Vol. 1 No. 4
第 4 期 Journal of China Agricultural University 1996

Goldstein 线搜索下共轭梯度法的全局收敛性①

李正锋② 陈 静 邓乃扬
(基础科学部)

摘 要 假设目标函数 $f(x)$ 在水平集上有下界且二次连续可微，证明了带 Goldstein 线搜索的共轭梯度法产生的搜索方向 d_k 是下降方向，并有 $\lim\limits_{k\to\infty}\inf\|\nabla f(x_k)\|=0$。作为一种特殊情形，得到了带 Goldstein 线搜索的 Fletcher-Reeves 方法的全局收敛性。

关键词 共轭梯度法；Fletcher-Reeves 方法；Goldstein 线搜索；全局收敛性

中图分类号 O221.2

Convergence Properties of Conjugate Gradient Methods With Goldstein Line Searches

Li Zhengfeng Chen Jing Deng Naiyang
(Department of Basic Sciences)

Abstract Assuming that the objective function $f(x)$ is bounded from below and is twice continuously differentiable on its level set, it is approved that the search directions generated by the conjugate gradient with Goldstein line searches are descent, and $\lim\limits_{k\to\infty}\inf\|\nabla f(x_k)\|=0$. As a special case, the global convergence property of the Fletcher-Reeves method with Goldstein line searches is derived.

Key words conjugate gradient method; Fletcher-Reeves method; Goldstein line search; global convergence

共轭梯度法是求解大规模无约束优化问题

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

的有效数值方法之一，其解为

$$x_{k+1}=x_k+\lambda_k d_k \tag{1}$$

$$d_k=\begin{cases}-g_k & k=1\\ -g_k+\beta_k d_{k-1} & k\geq 2\end{cases} \tag{2}$$

收稿日期：1995-09-26
①国家自然科学基金和北京市自然科学基金资助项目
②李正锋，北京清华东路 17 号中国农业大学(东校区)71 信箱，100083

其中：$\lambda_k>0$，为线搜索步长；d_k 为下降方向；$g_k=\nabla f(x_k)$；β_k 为一参数，通常选取 β_k 使得当目标函数 $f(x)$ 是二次函数且线搜索是精确线搜索时，该方法具有 n 步终止性。选取 β_k 的著名方法有 Fletcher-Reeves(FR)方法和 Polak-Ribière-Polyak(PRP)方法，选取的 β_k 分别为

$$\beta_k^{FR}=\frac{\|g_k\|^2}{\|g_{k-1}\|^2}$$

$$\beta_k^{PRP}=g_k^T(g_k-g_{k-1})/\|g_{k-1}\|^2$$

为了不失一般性，本文中只假定 β_k 满足

$$|\beta_k/\beta_k^{FR}|\leqslant 1 \tag{3}$$

共轭梯度法基于强 Wolfe 线搜索条件下的全局收敛性已有许多人研究过[1~4]，然而极少有人讨论共轭梯度法在其他线搜索条件下的收敛性质。我们知道，当目标函数的梯度用有限差分近似时，Goldstein 线搜索比 Wolfe 线搜索的函数值计算次数少，因此本文中研究共轭梯度法在 Goldstein 线搜索下的收敛性质是很有意义的。所谓强 Wolfe 搜索条件是指线搜索步长 λ_k 满足

$$f(x_k+\lambda_k d_k)-f(x_k)\leqslant\alpha\lambda_k g_k^T d_k \tag{4}$$

$$|g(x_k+\lambda_k d_k)^T d_k|\leqslant-\beta g_k^T d_k$$

其中 $0<\alpha<\beta<1/2$。Goldstein 线搜索要求步长 λ_k 满足

$$\alpha_2\lambda_k g_k^T d_k\leqslant f(x_{k+1})-f(x_k)\leqslant\alpha_1\lambda_k g_k^T d_k \tag{5}$$

其中 $0<\alpha_1<1/2<\alpha_2<1$。式(5)右端相当于条件(4)。

1 Goldstein 线搜索的性质

假定目标函数 $f(x)$ 满足假设 1。

假设 1

1) $f(x)$ 在水平集 $L=\{x\,|\,f(x)\leqslant f(x_1)\}$ 上有下界；

2) $f(x)$ 在 L 的一个邻域 N 内二次连续可微且其 Hessian 阵满足

$$\|\nabla^2 f(x)\|\leqslant M,\quad \forall x\in L \tag{6}$$

可以证明，如果 $f(x)$ 有下界且 $g_k^T d_k<0$，则满足式(5)的 λ_k 是存在的[5]。

定理 1 设 x_1 是使目标函数满足假设 1 的初始点，考虑形如(1)和(2)的任意点列，步长 λ_k 满足 Goldstein 条件式(5)，则

$$\sum_{k=1}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2}<+\infty \tag{7}$$

证明 由式(5)右端有

$$\sum_{k=1}^{\infty}|\lambda_k g_k^T d_k|<\infty \tag{8}$$

从泰勒公式和式(6)可得到

$$f(x_{k+1})-f(x_k)=\lambda_k g_k^T d_k+\int_0^1(1-\theta)\lambda_k^2 d_k^T\nabla^2 f(x_k+\theta\lambda_k d_k)d_k\mathrm{d}\theta\leqslant$$

$$\lambda_k g_k^T d_k+M\lambda_k^2\|d_k\|^2/2$$

注意到式(5)左端又有

$$\lambda_k g_k^{\mathrm{T}} d_k + M\lambda_k^2 \|d_k\|^2/2 \geqslant \alpha_2 \lambda_k g_k^{\mathrm{T}} d_k \tag{9}$$

利用式(9)可知

$$\lambda_k \geqslant \frac{2}{M} \frac{(\alpha_2 - 1) g_k^{\mathrm{T}} d_k}{\|d_k\|^2} \tag{10}$$

从式(8)和(10)可得(7)。

对于一般的 β_k,例如 β_k^{FR},容易举例证明 Goldstein 线搜索不能像 Wolfe 线搜索那样保证产生的搜索方向 d_k 是下降的。

由式(2)有

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

若 $g_k^{\mathrm{T}} d_{k-1} > 0$ 且 $\beta_k < 0$,则 $g_k^{\mathrm{T}} d_k < 0$;若 $g_k^{\mathrm{T}} d_{k-1} < 0$ 且 $\beta_k > 0$,仍有 $g_k^{\mathrm{T}} d_k < 0$。因此,总可以选取 β_k 使得式(2)在 Goldstein 线搜索下产生的搜索方向 d_k 总是下降方向。

2　全局收敛性

采用 Goldstein 线搜索的共轭梯度法的步骤如下。

算法 1

1)取初始点 x_1,令 $k=1$;

2)计算$\|g_k\|$,若 $g_k=0$,则停,否则继续;

3)按式(2)计算 d_k,并且选取参数 β_k 和步长 λ_k 使得式(3)和(5)成立;

4)用式(1)计算 x_{k+1},令 $k=k+1$,转 2)。

下面讨论这种算法的收敛性。首先给出一个引理。

引理 1[4]　设 $l(>0)$和 c 为常数,如果正项级数$\{a_i\}$对所有 k 都有

$$\sum_{i=1}^{k} a_i \geqslant lk + c$$

则

$$\sum_{i \geqslant 1} \frac{a_i^2}{i} = \infty$$

$$\sum_{k \geqslant 1} \frac{a_k^2}{\sum_{i=1}^{k} a_i} = \infty$$

定理 2　设 x_1 是使目标函数 $f(x)$满足假设 1 的初始点,如果存在常数 $M(>0)$使得

$$\|\nabla f(x)\| \leqslant M, \quad \forall x \in L \tag{11}$$

则对算法 1 产生的迭代点列$\{x_k\}$有

$$\liminf_{k \to \infty} \|g_k\| = 0$$

证明　由式(2)知,$k \geqslant 2$ 时有

$$d_k + g_k = \beta_k d_{k-1}$$

两边取平方模并移项得

$$\|d_k\|^2 = -\|g_k\|^2 - 2g_k^{\mathrm{T}} d_k + \beta_k^2 \|d_{k-1}\|^2 \tag{12}$$

令

$$t_k = \frac{\|d_k\|^2}{\|g_k\|^4}, \quad r_k = -\frac{g_k^{\mathrm{T}} d_k}{\|g_k\|^2}$$

式(12)和(3)表明

$$t_k \leqslant t_{k-1} - \frac{1}{\|g_k\|^2} - \frac{2r_k}{\|g_k\|^2}$$

注意到 $t_1 = 1/\|g_1\|^2, r_1 = 1$,则有

$$t_k \leqslant -\sum_{i=1}^{k} \frac{1}{\|g_i\|^2} + 2\sum_{i=1}^{k} \frac{|r_i|}{\|g_i\|^2} \tag{13}$$

现在用反证法证。假设

$$\liminf_{k\to\infty} \|g_k\| \neq 0$$

则存在常数 $\tau > 0$,使得

$$\|g_k\| \geqslant \tau \tag{14}$$

对所有 k 都成立,这时,在式(13)中应用式(11)和(14)得

$$t_k \leqslant -\frac{k}{M} + \frac{2}{\tau} \sum_{i=1}^{k} |r_i| \tag{15}$$

故

$$t_k \leqslant \frac{2}{\tau} \sum_{i=1}^{k} |r_i| \tag{16}$$

另一方面,因 $t_k \geqslant 0$,由式(15)又有

$$\sum_{i=1}^{k} |r_i| \geqslant \frac{\tau k}{2M} \tag{17}$$

据式(16)和(17)及引理 1 知

$$\sum_{k\geqslant 1} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} = \sum_{k\geqslant 1} \frac{r_k^2}{t_k} = \infty$$

这与定理 1 矛盾,故 $\liminf_{k\to\infty} \|g_k\| = 0$,定理得证。

令 $\beta_k = \beta_k^{\mathrm{FR}}$,式(3)仍成立,可得如下定理。

定理 3 假设条件与定理 2 相同,则 Fletcher-Reeves 算法产生的迭代点列 $\{x_k\}$ 满足

$$\liminf_{k\to\infty} \|g_k\| = 0$$

参 考 文 献

1 AL-Baali M. Descent properties and global convergence of the Fletcher-Reeves method with inexact line search. IMA J Numer Anal,1985,5:121～124

2 Gilbert J L, Nocedal J. Global convergence properties of conjugate gradient methods for optimization. SIAM J Optimization,1992,2(1):21～42

3 Liu Guanghui, Han Jiye, Yin Hongxia. Global convergence of the Fletcher-Reeves algorithm with inexact line search. Report of Institute of Applied Mathematics, Academia, Sinica. Beijing: Academia, Sinica, 1993

4 戴彧虹,袁亚湘. 广义 Wolfe 线搜索下 Fletcher-Reeves 方法的收敛性. 高等学校计算数学学报(待发表)

5 席少霖编. 非线性最优化方法. 北京:高等教育出版社,1992. 469 页